

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-031742

(43)Date of publication of application : 03.02.1998

(51)Int.Cl.

G06T 7/00

G01B 11/00

G06T 1/00

G06T 7/60

(21)Application number : 08-184917

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 15.07.1996

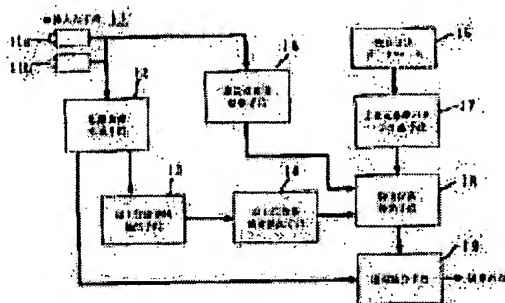
(72)Inventor : HASHIMOTO MANABU  
WASHIMI KAZUHIKO

## (54) IMAGE PROCESSOR AND OBJECT TRANSFER DEVICE

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To recognize a three-dimensional position of an object in a short time by an image memory of a small capacity and simple device constitution by generating a distance image of loaded a plurality of objects and extracting a top stage surface area from the distance image.

**SOLUTION:** The image of a recognition object is inputted to a distance image generation means 12 by an image input means 11 and the distance image is generated. Then, by a top stage surface area extraction means 13, the top stage surface area provided with a height equivalent to the object of a top stage is extracted. Since the object of the top stage surface area expressed by binary values looks like a rectangular pattern on the image, by detecting side edge parts crossed at the right angle, a candidate is extracted. An object candidate is positioned by using a template matching method by an object position detection means 17. By coarsely generating the distance image, extracting the candidate of the object from there and executing detailed position detection in the form of extremely limiting the range of pattern retrieval, the distance image generator is made small in the device scale.



### \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.\*\*\* shows the word which can not be translated.

3.In the drawings, any words are not translated.

## CLAIMS

[Claim(s)]

[Claim 1] A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, An object size database which stores size data of recognition objects, and a two-dimensional reference pattern creating means which generates a standard pattern image on an objective two-dimensional picture based on objective size data, A shade original image storing means which stores a shade original image inputted from a camera, An object position detecting means which detects an objective position using information on the above-mentioned two-dimensional reference pattern and the above-mentioned shade original image about a highest rung object candidate extracted according to the above-mentioned individual, An image processing device which is provided with an information integration means which unifies a detection result of an object position obtained by each above-mentioned means, and distance information of each object obtained by a depth map creating means, and outputs three dimensional position information on each object body.

[Claim 2] A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, An object size database which stores size data of recognition objects, and a two-dimensional reference pattern creating means which generates a standard pattern image on an objective two-dimensional picture based on objective size data, A shade original image storing means which stores a shade original image inputted from a camera, An object position detecting means which detects an objective position using information on a two-dimensional reference pattern and a shade original image about a highest rung object candidate extracted according to the above-mentioned individual, It has an information integration means which unifies a detection result of the above-mentioned object, and distance information of each object obtained by a depth map creating means, An object transfer equipment which is provided with an object transfer means which grasps and transfers an image processing device and an object which output three dimensional position information on each object body, and transfers an object based on an output signal of a described image processing unit.

[Claim 3] A random texture light projection means which floodlights a texture pattern random in a depth map creating means, The image processing device according to claim 1 comprising a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the first image input means and the second image input means which input a stereo image, and the above-mentioned first and the second image input means picturized.

[Claim 4] A random texture light projection means in which a depth map creating means floodlights a random texture pattern, The first image input means and second image input means that input a stereo image, The object transfer equipment according to claim 2 which is provided with an object transfer means which grasps and transfers an image processing device and an object which comprised a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the above-mentioned first and the second image input means picturized, and transfers an object based on an output signal of a described image processing unit.

[Claim 5] A low resolution depth map creating means in which a depth map creating means generates a depth map with coarse resolution, A level difference region extraction means to extract an objective level difference region part from a low resolution depth map, The image processing device according to claim 1 comprising a high resolution depth map creating means which generates a depth map of high resolution to a level difference region, and a depth map synthesizing means which compounds two depth maps from which resolution of the above-mentioned low resolution depth map and a high resolution depth map differs.

[Claim 6] A low resolution depth map creating means in which a depth map creating means

generates a depth map with coarse resolution, A level difference region extraction means to extract an objective level difference region part from a low resolution depth map, A high resolution depth map creating means which generates a depth map of high resolution to a level difference region, The object transfer equipment according to claim 2 which is provided with an object transfer means which grasps and transfers an image processing device and an object which comprised a depth map synthesizing means which compounds two depth maps from which resolution of the above-mentioned low resolution depth map and a high resolution depth map differs, and transfers an object based on an output signal of a described image processing unit. [Claim 7] A high-speed similarity calculation means by which a with means corresponding to a block searches for a correspondence image block on either side at high speed, The image processing device according to claim 2 which comprised a high reliability similarity calculation means to calculate similarity to high reliability about a pair of a block which is search results, and a corresponding block reliability judgment means will judge that have no corresponding block if the above-mentioned high reliability similarity is below a fixed threshold.

[Claim 8] A high-speed similarity calculation means by which a with means corresponding to a block searches for a correspondence image block on either side at high speed, A high reliability similarity calculation means to calculate similarity to high reliability about a pair of a block which is search results, The object transfer equipment according to claim 4 which will be provided with an object transfer means which grasps and transfers an image processing device and an object which comprised a corresponding block reliability judgment means judge that has no corresponding block if the above-mentioned high reliability similarity is below a fixed threshold, and transfers an object based on an output signal of a described image processing unit.

[Claim 9] Have three or more image input means, and a with means corresponding to a block chooses two or more kinds of picture pairs which chose an inner 2 \*\* picture of each picture acquired from the three or more above-mentioned image input means, The image processing device according to claim 1 or 3 performing with [ corresponding to a block ], unifying each search results with correspondence, and generating a depth map.

[Claim 10] Have three or more image input means, and a with means corresponding to a block chooses two or more kinds of picture pairs which chose an inner 2 \*\* picture of each picture acquired from the three or more above-mentioned image input means, The object transfer equipment according to claim 2 or 4 which performs with [ corresponding to a block ], is provided with an object transfer means which grasps and transfers an image processing device and an object which unify each search results with correspondence and generate a depth map, and transfers an object based on an output signal of a described image processing unit.

[Claim 11] A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, An object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from an output of a highest rung surface area extraction means, An object size database which stores size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of an object candidate combination hypothesis which the above-mentioned hypothesis generating means generated using information included in a shade original image, An image processing device which recognizes based on a hypothesis with the highest evaluation value among two or more hypotheses, is provided with an information integration means which unifies this recognized picture and distance information of each object obtained by a depth map creating means, and outputs three dimensional position information on each object body.

[Claim 12] A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, An object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from an output of a highest rung surface area extraction means, An object size database which stores size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of an object

candidate combination hypothesis which the above-mentioned hypothesis generating means generated using information included in a shade original image, Recognize based on a hypothesis with the highest evaluation value among two or more hypotheses, and it has an information integration means which unifies this recognized picture and distance information of each object obtained by a depth map creating means, An object transfer equipment which is provided with a transfer means which grasps and transfers an image processing device and an object which output three dimensional position information on each object body, and transfers an object based on an output signal of a described image processing unit.

[Claim 13]An object arrangement reference database which stores data used as a standard of object arrangement, A picture which a hypothetical validation means to verify the validity of an object candidate combination hypothesis which a hypothesis generating means generated has recognized, The image processing device according to claim 11 having an object arrangement judging means which judges whether it is right arrangement as compared with the above-mentioned criterion data, and a warning generation means which warns a worker of a result of a judgment when a recognized picture does not consistent with criterion data.

[Claim 14]The object transfer equipment according to claim 12 which is provided with a transfer means which grasps and transfers an image processing device and an object characterized by comprising the following, and transfers an object based on an output signal of a described image processing unit.

An object arrangement reference database which stores data used as a standard of object arrangement.

A picture which a hypothetical validation means to verify the validity of an object candidate combination hypothesis which a hypothesis generating means generated has recognized.

An object arrangement judging means which judges whether it is right arrangement as compared with the above-mentioned criterion data.

A warning generation means which warns a worker of a result of a judgment when a recognized picture does not consistent with criterion data.

[Claim 15]The image processing device according to claim 1 which is provided with the following and characterized by operating so that the above-mentioned actual image reference pattern may be used in subsequent recognition operation.

An actual image reference pattern logging means which starts an imaging range which detects a position of the target object from an object size database using an artificial reference pattern generated automatically, and is equivalent to the above-mentioned object body in a detection position.

An actual image reference pattern storing means which stores a cut-down actual image reference pattern.

[Claim 16]A position of the target object is detected from an object size database using an artificial reference pattern generated automatically, An actual image reference pattern logging means which starts an imaging range which is equivalent to the above-mentioned object body in a detection position, It has an actual image reference pattern storing means which stores a cut-down actual image reference pattern, The object transfer equipment according to claim 2 which is provided with a transfer means which grasps and transfers an image processing device and an object which operate so that the above-mentioned actual image reference pattern may be used in subsequent recognition operation, and transfers an object based on an output signal of a described image processing unit.

[Claim 17]A depth map creating means which measures distance distribution from an image input means of two or more tiered object bodies, Based on information on a depth map stored in the first depth map storing means that stores a generated depth map, and this first depth map creating means, at the time of an initial bundle which calculates a position at the time of an objective bundle A position calculation means, The second depth map storing means that stores a depth map of the remaining cargo after evacuating an object to a position temporarily, A depth map comparison means to compare a depth map stored in a depth map storing means of the



above first with a depth map stored in a depth map storing means of the above second, An image processing device having a position-correction-quantity detection means to determine a gripping position of the evacuation above-mentioned object based on a comparison result of this depth map comparison means, and to detect a difference of this gripping position and the above-mentioned initial gripping position as position correction value.

[Claim 18] A depth map creating means which measures distance distribution from an image input means of two or more tiered object bodies, The first depth map storing means that stores a generated depth map, and an object transportation device which grasps an object and moves, At the time of an initial bundle which calculates a position based on information on the above-mentioned depth map at the time of an objective bundle, a position calculation means, The second depth map storing means that stores a depth map of the remaining cargo after grasping an object in the above-mentioned gripping position and making it contrast with a position temporarily by the above-mentioned object transportation device, A depth map comparison means to compare a depth map stored in a depth map storing means of the above first with a depth map stored in a depth map storing means of the above second, Based on an output of an image processing device provided with an image processing device which has a position-correction-quantity detection means to determine a gripping position of an object which evacuated [ above-mentioned ] using a comparison result of the above-mentioned depth map, and to detect a difference with the above-mentioned initial gripping position as position correction value, and an objective transfer equipment, An object transfer equipment which transfers an object by an object transfer means.

---

[Translation done.]

**\*NOTICES\***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**DETAILED DESCRIPTION**

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the image processing device which recognizes the position of the load into which the robot which transfers a cargo automatically, and the industrial machine were loaded.

[0002]

[Description of the Prior Art] There are some which were indicated by JP,H6-249631,A as an example of the image processing device of a means to transfer a cargo automatically. The flow chart with which drawing 27 shows the approximate account figure of the procedure of the image measuring of the invention and drawing 28 operation, and drawing 29 are the figures showing the principle of spatial-coding which used pattern light.

[0003] This image processing device makes applicable to Measurement Division carton box W loaded on the palette 1.

As shown in drawing 29, when it becomes the slanting upper part of the palette 1 from the floodlight 2 and the camera 3 arranged in the central part upper part of the palette 1 and the floodlight 2 floodlights various kinds of pattern light serially, The slit pattern light which divides

measuring space into the wedge-shaped measurement regions  $r_0$ - $r_n$  defined mutually is floodlighted.

A slit pattern floodlights three pattern light of the monochrome binary pattern of A of drawing 29, B, and C, for example.

As for white, light hits and a portion and black are portions upon which light does not strike. Such a pattern can be made, for example using a dot-matrix electrical-and-electric-equipment shutter like a liquid crystal shutter.

[0004] If the state where light has hit is expressed as "1" and the state where it has not hit is expressed as "0", when the pattern A is floodlighted, in drawing 29, it will be set to 1 and 0 from before by halves toward back. In the pattern B, it is set to 1, 0, 1, and 0 from before every  $[4 / 1/]$  at order. In the pattern C, it is set to 1, 0, 1, 0, 1, 0, 1, and 0 every  $[8 / 1/]$ .

[0005] If these three patterns are floodlighted and a picture is picturized, respectively, it can be made to correspond to one of the data coded by three bit codes of "000" to "111" about each pixel. Such a code is called a spatial code. A wedge-shaped field corresponds to a certain spatial code at three-dimensional Jo Sorama. A code will be assigned to the imaging range on the objective surface when the object W exists there. On the other hand, since the picture picturized with the camera shows whether it is equivalent to which wedge-shaped field if the spatial code of a certain pixel on a picture is known, the principle of triangulation will show the height of an object surface, i.e., the distance from a camera. Although the kind of pattern was written with three kinds, A, B, and C, and the whole picture was divided into eight wedge-shaped fields in this explanation, if the kind of pattern is made into eight kinds, the number of wedge-shaped fields can be set to 256, and can raise the accuracy of distance information more precisely. That is, the distance from the camera of each pixel on a picture is measurable by generating a spatial code picture.

[0006] Drawing 27 is explained according to a flow chart. If a system is started by step ST01, a spatial code picture will be generated in step ST02. A spatial code picture as shown in the figure (D) is acquired by the upper surface of the cargo in the state of drawing 27 (A), a picture is horizontally scanned on the right from the left, and if a field with data with the largest spatial code is extracted, the picture a highest rung object carries out [ a picture ] upper surface partial load correspondence as shown in the figure (E) will be acquired. In step ST04, grouping of the highest rung object area is carried out. When the spatial code is close within the limits of predetermined between the pixels which adjoin in a spatial code picture, these fields are judged to be the same groups, or grouping is carried out based on change of a grooved portion or the spatial code for liking Mabe produced between each cargo upper limb. The three-dimensional position of each object is measured by step ST05 which each object by which grouping was carried out follows. A number of three-dimensional coordinates of the portion equivalent to the verge portion of the object by which grouping was carried out are sampled, and an objective centroid position, a posture, and height information are acquired based on them by step ST06. Processing is ended by step ST07. Thus, every one cargo is transferable by a robot etc. using the position data of the obtained cargo.

[0007] In this image processing device, all information required for recognition of a cargo is acquired from a spatial code picture. For example, it is necessary to generate the depth map of the resolution 256x256 or resolution as high as 512x512 pixels. Since the highly precise pattern floodlight which has a mechanism to which a floodlighting pattern is changed serially is used, two or more floodlighting patterns of each generated in the time series are made to correspond and a picture is inputted, a mass image memory is needed.

[0008] As other conventional examples, Sato published by Institute of Electronics, Information and Communication Engineers paper magazine Vol.J71-D.No.7p1240 - 1257 and the contents described in the paper "a liquid crystal range finder .. high-speed depth map instrumentation system by a liquid crystal shutter .." of Inokuchi are introduced. This art is related with the depth map acquisition system by the spatial code-ized method used with the art shown in above-mentioned drawing 27 - drawing 29, and shows drawing 30 that principle. The slit pattern by reflected binary code is serially projected on an object using the liquid crystal shutter 41, the picture corresponding to each pattern is picturized with the camera 3, and a spatial code picture

is generated. As a pattern, a pattern of seven sheets like drawing 4 is used. The principle of Measurement Division is the same as that of what is used by the conventional technology of above-mentioned drawing 27 - drawing 29.

[0009]As another conventional technology, the "stereo-correspondences searching device" indicated to JP,H7-103734,A is explained. Drawing 31 is an explanatory view of the hierarchical block matching (heuristics of condensation and rarefaction) of this invention.

Drawing 32 is a flow chart of operation of the invention.

[0010]Stereo vision is the technique of calculating the distance of a camera and a target point from the principle of triangulation by measuring the azimuth difference corresponding to the same point on two subjects of a picture inputted from the image input means arranged at several different positions. Processing which changes into hierarchical more low resolution the picture which two pictures which the picture was inputted in ST11 of the flow chart of drawing 32, and were inputted by ST12 were stored in the image memory, respectively, and was stored in ST13 is performed, and a hierarchy picture is generation \*\*\*\*. Drawing 31 expresses the correspondence relation between the Dth hierarchy's picture, and the D+1st hierarchies' picture in \*\* type. In this example, 1/4 field of the Dth hierarchy's picture is made into the D+1st hierarchies' picture. In [ a picture is divided into a small block in step ST14, and ] step ST905, Initial search positions are calculated by an initial search-positions computer, the corresponding points of the block in other pictures are calculated by pattern matching, and, as for each block, azimuth difference is calculated by an azimuth difference computer in step ST17 using the position information on the corresponding points between two pictures by step ST16. Thus, based on the obtained parallax picture, the picture of single step high resolution is used and azimuth difference is calculated again. This processing is repeated and corresponding point searching to the picture of the same peak resolution as an inputted image is given to the deed last target.

[0011]

[Problem(s) to be Solved by the Invention]In the image processing device shown in above-mentioned drawing 27 - drawing 29. All information required for recognition of a cargo is acquired from the spatial code picture, For example, it is necessary to generate the depth map of the resolution 256x256 or resolution as high as 512x512 pixels. A highly precise pattern floodlight is required and moreover what has a mechanism to which a floodlighting pattern is changed serially is required for it, It became a large-scale, expensive device, and two or more floodlighting patterns of each generated in the time series needed to be made to correspond further, the picture needed to be inputted, the mass image memory was also needed, and there was a problem used as a large-scale, expensive device. There was the total of the image input time which it takes, and many or the problem of cutting.

[0012]also in the art shown in above-mentioned drawing 30, since the cargo had been recognized only based on a spatial code picture, like the case of drawing 27 - drawing 29, the big image memory was also needed and there was a problem used as a large-scale, expensive device.

[0013]Drawing 27 - drawing 29, and the image processing device of drawing 30, In order to acquire all information required for a cargo from a spatial code picture, when an object and a neighboring matter object to grasp in a cargo touch densely, Namely, the thing in which a boundary with the object which adjoins in a carton box with the small amount of camfering of a ridgeline does not appear as a level difference clearly in three dimensions, Or since it was close with an adjoining object when it is pouches like the cement bag in which soft contents were stored, there was a problem that a boundary was indefinite in three dimensions and it was difficult to separate each object individually and to make a transporting means grasp.

[0014]Although the image processing device shown in drawing 31 and drawing 32 is searched for a stereo-correspondences point at high speed and it aims at gaining a depth map in short time with it, Since order searched in the small block unit which the subject in a picture was not caused how but was always uniformly set up on the picture from the picture of rougher resolution to the picture of dense resolution, when the part which needs high resolution information was found beforehand, there was a problem that processing time started for a long

time conversely.

[0015] This invention is made in order to solve the above-mentioned problem, and it is a thing. It is providing the object transfer equipment equipped with the image processing device which does not need the purpose but can grasp an objective three-dimensional position by an easy equipment configuration in a short time, and this image processing device.

[0016]

[Means for Solving the Problem] A depth map creating means which generates a depth map of two or more objects into which an image processing device concerning Claim 1 of this invention was loaded, A highest rung surface area extraction means to extract highest rung surface area of an object loaded from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, An object size database which stores size data of recognition objects, and a two-dimensional reference pattern creating means which generates a standard pattern image on an objective two-dimensional picture based on objective size data, A shade original image storing means which stores a shade original image inputted from a camera, An object position detecting means which detects an objective position using information on a two-dimensional reference pattern and a shade original image about a highest rung object candidate extracted individually, It has an information integration means which unifies a detection result of an object position obtained by each means, and distance information of each object obtained by a depth map creating means.

[0017] A depth map creating means which generates a depth map of two or more objects into which an object transfer equipment concerning Claim 2 of this invention was loaded, A highest rung surface area extraction means to extract highest rung surface area of an object loaded from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, An object size database which stores size data of recognition objects, and a two-dimensional reference pattern creating means which generates a standard pattern image on an objective two-dimensional picture based on objective size data, A shade original image storing means which stores a shade original image inputted from a camera, An object position detecting means which detects an objective position using information on a two-dimensional reference pattern and a shade original image about a highest rung object candidate extracted individually, It has an object transfer means which grasps and transfers an image processing device and an object provided with an information integration means which unifies a detection result of an object position obtained by each means, and distance information of each object obtained by a depth map creating means.

[0018] A random texture light projection means which floodlights a texture pattern random [ image processing device / concerning Claim 3 of this invention ] in a depth map creating means, It comprises a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the first image input means and the second image input means which input a stereo image, and the first and the second image input means picturized.

[0019] An object transfer equipment concerning Claim 4 of this invention is characterized by comprising:

A random texture light projection means in which a depth map creating means floodlights a random texture pattern.

The first image input means and second image input means that input a stereo image.

An image processing device which comprised a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the first and the second image input means picturized.

An object transfer means which grasps and transfers an object.

[0020] A low resolution depth map creating means in which, as for an image processing device concerning Claim 5 of this invention, a depth map creating means generates a depth map with coarse resolution, A level difference region extraction means to extract an objective level difference region part from a low resolution depth map, It comprises a high resolution depth map

creating means which generates a depth map of high resolution to a level difference region, and a depth map synthesizing means which compounds two depth maps from which resolution of a low resolution depth map and a high resolution depth map differs.

[0021]An object transfer equipment concerning Claim 6 of this invention is characterized by comprising:

A low resolution depth map creating means in which a depth map creating means generates a depth map with coarse resolution.

A level difference region extraction means to extract an objective level difference region part from a low resolution depth map.

A high resolution depth map creating means which generates a depth map of high resolution to a level difference region.

An image processing device which comprised a depth map synthesizing means which compounds two depth maps from which resolution of a low resolution depth map and a high resolution depth map differs, and an object transfer means which grasps and transfers an object.

[0022]A high-speed similarity calculation means by which, as for an image processing device concerning Claim 7 of this invention, a with means corresponding to a block searches for a correspondence image block on either side at high speed. It comprises a high reliability similarity calculation means to calculate similarity to high reliability about a pair of a block which is search results, and a corresponding block reliability judgment means will judge that have no corresponding block if high reliability similarity is below a fixed threshold.

[0023]An object transfer equipment concerning Claim 8 of this invention is characterized by comprising:

A high-speed similarity calculation means by which a with means corresponding to a block searches for a correspondence image block on either side at high speed.

A high-reliability similarity calculation means to calculate similarity to high reliability about a pair of a block which is search results.

An image processing device which comprised a corresponding block reliability judgment means will judge that has no corresponding block if high reliability similarity is below a fixed threshold.

An object transfer means which grasps and transfers an object.

[0024]An image processing device concerning Claim 9 of this invention has three or more image input means, A with means corresponding to a block chooses two or more kinds of picture pairs which chose an inner 2 \*\* picture of each picture acquired from three or more image input means, performs with [ corresponding to a block ], unifies each search results with correspondence, and generates a depth map.

[0025]An object transfer equipment concerning Claim 10 of this invention is characterized by comprising:

An image processing device which has three or more image input means, chooses two or more kinds of picture pairs which chose an inner 2 \*\* picture of each picture by which a with means corresponding to a block is obtained from the three or more above-mentioned image input means, performs with [ corresponding to a block ], unifies each search results with correspondence, and generates a depth map.

An object transfer means which grasps and transfers an object.

[0026]An image processing device concerning Claim 11 of this invention, A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, An object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from an output of a highest rung surface area extraction means, An object size database which stores size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of an object candidate combination hypothesis which the above-mentioned hypothesis generating means generated using information included in a shade

original image, It recognizes based on a hypothesis with the highest evaluation value among two or more hypotheses, and has an information integration means which unifies this recognized picture and distance information of each object obtained by a depth map creating means.

[0027]An object transfer equipment concerning Claim 12 of this invention, A depth map creating means which generates a depth map of two or more loaded objects, and a highest rung surface area extraction means to extract highest rung surface area of an object located in the highest rung from a depth map, An object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from an output of a highest rung surface area extraction means, An object size database which stores size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of an object candidate combination hypothesis which the above-mentioned hypothesis generating means generated using information included in a shade original image, Recognize based on a hypothesis with the highest evaluation value among two or more hypotheses, and it has an information integration means which unifies this recognized picture and distance information of each object obtained by a depth map creating means, It has an image processing device which outputs three dimensional position information on each object body, and a transfer means which grasps and transfers an object.

[0028]An image processing device concerning Claim 13 of this invention, An object arrangement reference database which stores data used as a standard of object arrangement, A picture which a hypothetical validation means to verify the validity of an object candidate combination hypothesis which a hypothesis generating means generated has recognized, It has an object arrangement judging means which judges whether it is right arrangement as compared with criterion data, and a warning generation means which warns a worker of a result of a judgment when a recognized picture does not consistent with criterion data.

[0029]An object transfer equipment concerning Claim 14 of this invention is characterized by comprising:

An object arrangement reference database which stores data used as a standard of object arrangement.

A picture which a hypothetical validation means to verify the validity of an object candidate combination hypothesis which a hypothesis generating means generated has recognized.

An object arrangement judging means which judges whether it is right arrangement as compared with the above-mentioned criterion data.

An image processing device which has a warning generation means which warns a worker of a result of a judgment when a recognized picture does not consistent with criterion data, and a transfer means which grasps and transfers an object.

[0030]An image processing device concerning Claim 15 of this invention, A position of the target object is detected from an object size database which stores data used as a standard of object arrangement using an artificial reference pattern generated automatically, An actual image reference pattern logging means which starts an imaging range which is equivalent to said object body in a detection position, It has an actual image reference pattern storing means which stores a cut-down actual image reference pattern, and it constitutes so that said actual image reference pattern may be used in subsequent recognition operation and it may operate.

[0031]An object transfer equipment concerning Claim 16 of this invention, A position of the target object is detected from an object size database which stores data used as a standard of object arrangement using an artificial reference pattern generated automatically, An actual image reference pattern logging means which starts an imaging range which is equivalent to an object body in a detection position, It has an actual image reference pattern storing means which stores a cut-down actual image reference pattern, and has an image processing device constituted so that an actual image reference pattern might be used in subsequent recognition operation and it might operate, and a transfer means which grasps and transfers an object.

[0032]An image processing device concerning Claim 17 of this invention, A depth map creating means which measures distance distribution from an image input means of two or more tiered object bodies, Based on information on a depth map stored in the first depth map storing means



that stores a generated depth map, and this first depth map creating means, at the time of an initial bundle which calculates a position at the time of an objective bundle A position calculation means, The second depth map storing means that stores a depth map of the remaining cargo after evacuating an object to a position temporarily, A depth map comparison means to compare a depth map stored in the first depth map storing means with a depth map stored in the second depth map storing means, A gripping position of an object evacuated based on a comparison result of this depth map comparison means is determined, and it has a position-correction-quantity detection means to detect a difference of this gripping position and an initial gripping position as position correction value.

[0033]An object transfer equipment concerning Claim 18 of this invention, A depth map creating means which measures distance distribution from an image input means of two or more tiered object bodies, The first depth map storing means that stores a generated depth map, and an object transportation device which grasps an object and moves, At the time of an initial bundle which calculates a position based on information on the above-mentioned depth map at the time of an objective bundle, a position calculation means, The second depth map storing means that stores a depth map of the remaining cargo after grasping an object in a gripping position and making it contrast with a position temporarily by an object transfer means, A depth map comparison means to compare a depth map stored in the first depth map storing means with a depth map stored in the second depth map storing means, It has an image processing device which has a position-correction-quantity detection means to determine a gripping position of an object evacuated using a comparison result of a depth map, and to detect a difference with an initial gripping position as position correction value, and a transfer equipment which grasps and transfers an object.

[0034]

[Embodiment of the Invention]

Drawing 2 and the picture of each generated stage are shown for the flow chart which shows drawing 1 and a flow of operation for the block diagram showing the composition of embodiment 1. embodiment 1. in drawing 3. Here, an object is explained as a state where the thing of identical shape and the same size is stacked, supposing polyhedral form [, for example like a carton box ] whose object which should be recognized that explanation of operation is easy to be understood is.

[0035]The image input means which comprised two cameras which 11 set the predetermined interval and have been arranged in drawing 1, The depth map creating means which generates a depth map from the picture in which the image input means 11 picturized 12, A highest rung surface area extraction means by which 13 extracts a highest rung side from a depth map, and 14 A highest rung object candidate extraction means, The shade image storing means in which 15 stores an original image, and 16 An object size database, The two-dimensional reference pattern creating means which generates the reference pattern as a two-dimensional pattern automatically based on the information for which 17 was stored in the object size database 16, The object position detecting means from which 18 detects a two-dimensional object position, and 19 are information integration means which unify the distance information of the object position detected by the object position detecting means 18, and a depth map.

[0036]The original image in which the image input means 11 picturized (a) of drawing 3, the highest rung object area where the highest rung surface area extraction means 13 extracted (b), The template of the reference pattern in which (c) was generated by the two-dimensional reference pattern creating means using object size data, the mimetic diagram showing one of the candidates by whom (d) was extracted from the highest rung object area, and (e) show a positioning situation by the law by template matching.

[0037]By step ST101, starting of a device will input into the depth map creating means 12 the picture of the recognition object shown in drawing 3 (a) by the image input means 11. A depth map is generated by the depth map creating means 12 in step ST102. As a creating means of a depth map, there is the section coding method or stereo-correspondences heuristics. A highest rung object area with the height which is equivalent to the object of the highest rung shown in drawing 3 (b) by the highest rung object area extraction means 13 by step ST103 is extracted.



The original image photoed by the image input means 11 by step ST104 is stored in the shade image storing means 15, and in step ST105 by the two-dimensional reference pattern creating means 17. The reference pattern as a two-dimensional pattern is automatically generated based on the information stored in the object size database 16. This reference pattern is called a template, and as shown in drawing 3 (C), the template is an outline plate expressing an objective contour part.

[0038]In step ST106, a candidate is extracted from the object of the highest rung object area expressed with the binary being visible to a rectangular pattern on a picture by detecting the peripheral area at which it crosses right-angled. Since the resolution of a depth map is coarse as above-mentioned, a candidate's position is inaccurate. In step ST107, said object candidate is positioned by the object position detecting means 17, said two-dimensional reference pattern is used for positioning, and processing by the template-matching method is made. However, since the rough position is already detected in the stage of candidate extraction, a candidate detection position to the processing which searches only the neighborhood very much is enough as matching. Therefore, template matching can also be performed at high speed. The situation of positioning is shown in drawing 3 (e). If an object position is detected, the two-dimensional object position information will be unified with the distance information by said depth map creating means 12 by the information integration means 19, and will be outputted as a result with final it.

[0039]In step ST108, it judges whether all the objects were detected, if it has not detected, it performs again from step ST106 among processing flows, and a series of processing step ST106 and ST107 are carried out about another object candidate. If all the objects are detected, recognition processing will be terminated by step ST109. It is performing the range of that pattern searching in the form which generated the depth map coarsely by this invention, extracted the objective candidate based on that result as above, and limited dramatically template matching whose detailed detecting position is a two-dimensional general-purpose processing method. The device scale of a depth map generating device can be made small.

[0040]It may be the method that resolution is made coarse and a spatial code-sized method is used as a depth map creating means, or a method like stereo vision or the optical cutting method by slit light scan, and what is necessary is just a device with which the depth map of coarse resolution may be generated.

[0041]Although the example of positioning by template matching based on profiling information was shown, the method of considering an objective outline to be linear combination similarly, and performing alignment for every neighborhood as a straight line may be used. The contents of the object size database may be used and the usual two-dimensional template matching may be used. Although it did not show clearly, especially concerning the data storage form used inside template matching processing, the method of enumerating and storing the position of the edge of a contour part other than the method of storing as a two-dimensional picture in an inside in one dimension may be used.

[0042]When extracting the candidate of a highest rung object, the technique of using the corner part from the highest rung surface area information acquired from the depth map was shown, but. In addition, for example, the same effect is done so even if it applies the method of detecting a position coarsely by the method of template matching of a binary using the template as a depth map of the binary prepared separately, using the depth map of the above-mentioned binary.

[0043]The operation flow chart of drawing 2 may be an example, and as long as the input/output relation of each means shown in drawing 1 is proper, and is clear and it has it, it may be another processing flow. For example, as for the step of depth map generation and a shade original image input, whichever may be performed first.

[0044]Although the above was the explanation which assumed the polyhedron, i.e., the single variety object of a box-like object, as an object body which should be recognized, even if it is a saccate object like a cement bag and a rice bag, image recognition can be realized with the same composition, for example, without changing the composition of this invention in any way. Next, about a situation in case a bag-like object object recognizes, the difference with the case in the

above-mentioned case shape object is explained.

[0045]In step ST105 of drawing 2, a two-dimensional reference pattern is automatically generated by the two-dimensional reference pattern creating means 16. This is equivalent to the template generation in template matching for positioning an object. In the case of a bag-like object, it is insufficient unlike the case shape object -- the verge changes -- just to have taken out the contour part of a rectangle like drawing 3 as said reference pattern, since shape was unstable. In this case, the partial template which expressed only the objective corner part with the outline, or the straight-line contour part template expressing an objective side part is generated automatically. In any case, it is generated with reference to the contents of the object size database 15 which stored the size of the bag-like object.

[0046]In the case of a bag-like object, in object candidate extraction of step ST106, it is difficult to extract a candidate using the information on an edge etc. which were described in the top and at which it crosses right-angled. In this case, consider that the picture as which the highest rung object area was expressed is a binary format image, and the object template of a binary is automatically generated from the information about the bag-like object model stored in the object size database 15. Rough candidate extraction is attained by carrying out binary template matching of these coarsely.

[0047]In step ST107, although an objective position is detected, the outline template of the corner part of the above objects or a rectilinear side edge portion is matched with the edge image obtained from the inputted image as a partial template here. An objective position can be operated to the same grade as the recognition of as opposed to a case shape object also about an object saccate as mentioned above judged and calculated synthetically from the position which two or more partial templates matched as a result of matching.

[0048]Although the situation where the object of single variety was stacked was assumed in embodiment 1. explained above, The size data of two or more objects is stored in the object size database 16. It is carrying out operation it is supposed that it is an object which generates automatically two or more two-dimensional reference patterns, and makes said two or more reference patterns of all match in the object position detecting means 17, and in which a template with the highest similarity exists. It can be made to operate accurately also to the situation where two or more objects of a different kind are stacked.

[0049]By building the image processing device of the above-mentioned composition into an object transfer equipment, it becomes an object transfer equipment which does not need a highly precise pattern floodlight and a mass image memory, and can also perform image input time in a short time and which operates correctly.

[0050]The flow chart which shows drawing 4 and a flow of operation for the block diagram showing the composition of embodiment 2. embodiment 2. is shown in drawing 5. It is an explanatory view in which drawing 6 shows the explanatory view of a random dot pattern exposure, and drawing 7 shows the method of the with processing corresponding to a block from a stereo image. This embodiment 2. carries out the portion of the depth map creating means of embodiment 1. with stereo correspondences by random texture pattern floodlighting, and generates a depth map, this point is different, and other composition is the same as embodiment 1.

[0051]In drawing 4, the random texture pattern light projection means in which 20 floodlights a random dot pattern, and 21 are image input means, and comprise the first image input means 21a and the second image input means 21b. 22 is a with means corresponding to a stereo image block. Other highest rung imaging range extraction means 13, highest rung object candidate extraction means 14, shade original image storing means 15, object size databases 16, two-dimensional reference pattern creating means 17, object position detecting means 18, and information integration means 19 are the same composition as embodiment 1.

[0052]According to the flow chart of drawing 5, operation is explained about this embodiment 2. below. If a power supply is turned on by step ST201, the power supply of the random texture pattern light projection means 20 will be turned on, and a random dot pattern will be projected to a recognition object thing. A random texture pattern light projection means floodlights a random dot pattern with the pattern floodlight 30 to the subject which should be recognized, as shown in

drawing 6 (a). A pattern floodlight is installed downward above objective like drawing 6, and projects a random dot as shown in drawing 6 (b) on a recognition object. In embodiment 2., although it is a carton box, the object body which should be recognized can be recognized, for example, even if it is a saccate object like a cement bag. A random dot is projected on the upper surface of a subject, and the stereo image of the picture of the couple picturized by the first image input means 21a and second image input means 21b is inputted in step ST203. A predetermined distance makes an optic axis and this stereo image is picturized with two cameras, the same resolution and the same focal distance.

[0053]Next, a random dot pattern is explained. The random dot shown in drawing 6 (b) comprises resolution of 128x128 at this embodiment 2., by the computer, each pixel is a square, white or a black value is assigned for every pixel in uniform random numbers, and the ratio of white and a black pixel number has become about 1:1. The mechanism in which the focus of a projection pattern is adjusted is formed in the pattern floodlight 30, and a focus is adjusted so that it may focus by an object surface based on the data, when a rough distance of a floodlight and an object body is known. When distance is not known at all beforehand, it operates so that a focus may suit in the distance adjusted a priori.

[0054]When the input of a stereo image pair finishes, the light projection means 20 of a random texture is turned off by step ST204. Next, a stereo image is processed for a stereo image by the with stage 22 corresponding to a block by step ST205. Drawing 7 (a) and (b) is a picture of the couple of the stereo image which photoed the object body. Suppose that drawing 7 (a) is called a left image and drawing 7 (b) is called a right image. Block matching is performed from dividing a right image into a lattice-like small block, as shown in a figure. Although the above-mentioned random dot pattern is actually projected on the left and a right image, in order [ for which the block bR which it is taking notice of among two or more small blocks divided and generated now is called a noticing block ] to simplify a figure, the pattern is not drawn here. It searches for the position on the left image corresponding to this noticing block bR. It looks for the pattern which has the same pattern as bR in the search range set up separately. Search by accumulation difference absolute value minimization is adopted as the method of search. This piles up the inside of the search range on a left image one by one, moving the noticing block bR, and the similarity makes it corresponding points with the position which becomes the best. Under the present circumstances, an accumulation difference absolute value is used as evaluation of similarity. For example, in a noticing block, when bR (i, j) and a left image are set to bL (i, j) and block size is made into NxN, accumulation difference absolute value  $S_{SAD}(dx, dy)$  in the position (dx, dy) on a left image is expressed with (the formula 1). In this embodiment 2., it is N= 16.

[0055]

[Equation 1]

$$S_{SAD}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \cdots \cdots \cdots (式1)$$

[0056] $S_{SAD}$  for which it asked by (the formula 1) -- the minimum -- becoming (dx, dy) -- it means that the block bL on the left image corresponding to the noticing block bR was found by detecting The stereo method is a method of making the difference into azimuth difference, being a principle of triangulation and acquiring the distance from a camera to an object from the position during the block acquired by carrying out in this way. With processing corresponding to a series of blocks explained above is repeatedly performed to all the small blocks on a right image, it searches for the position on a left image corresponding for every block, and a depth map is generated.

[0057]It is the same as embodiment 1. after this. In step ST207 which the field which is equivalent to a highest rung object face from a depth map by the highest rung surface area extraction means 13 is extracted in step ST206, and continues. The shade original image picturized by the first image input means 21a or the second image input means 21b is stored in the shade original image storing means 15. The random dot pattern is not projected with the thing natural at this time. In step ST208, a two-dimensional reference pattern is automatically

generated by the two-dimensional reference pattern creating means 16 using the size data of the recognition objects stored in the object size database 16.

[0058] In step ST209, a picture from which the above-mentioned highest rung surface area was extracted is divided into two, a highest rung side and the other field, and is binary-ized, and one of the highest rung object candidates is extracted by the highest rung object candidate extraction means 14. In step ST210, a precise position of candidate objects is measured and detected by the object position detecting means 18 [near / which was extracted / one object candidate's position] using shade original image information stored in the above-mentioned two-dimensional reference pattern and the object position detecting means 15. At Step 211, a two-dimensional position of determined candidate objects, Until it judges that it was unified with said depth map information by the information integration means 19, and all the objects have been recognized by step ST212 by it, Processings from step ST209 to ST211 are repeated, and if judgment of step ST212 is judged that all the objects have recognized, it will be eventually outputted as objective three-dimensional information.

[0059] If it is judged that all the objects have been recognized by processings from step ST209 to ST211, processing will be ended by step ST213. What an objective position became final and conclusive one by one based on a binary-ized depth map from which the above-mentioned highest rung surface area was extracted is eliminated, and a judgment of whether all the objects have been recognized is performed by investigating whether highest rung surface area which moreover remains on a picture exists.

[0060] As explained above, in this embodiment 2. Since a random dot pattern is floodlighted to recognition objects, even if it is a plain carton box in which a pattern does not exist, for example in the surface, it is possible to obtain distance data in surface each of that position -- an object surface -- a position of an address label etc. -- it is possible to obtain a stereo image which is not influenced by unfixed texture. Conducts matching for stereo-correspondences attachment can also apply simple block matching.

[0061] Although a 128x128-pixel random dot pattern was projected to a camera view as resolution of a camera in this embodiment 2., Although it was still finer, a coarse dot pattern could be projected depending on size of a recognition object and size of 16x16 pixels was adopted as block size of block matching, 8x8 pixels, 32x32 pixels, etc. may be sufficient. Shape of a block may not necessarily be a square, either. Although a search range at the time of matching was set as a rectangle by drawing 7, it is still the same as general stereo correspondences that the reliability of matching can be raised or processing time can be shortened of it by making it change variously according to a situation.

[0062] In this embodiment 2., although a monochrome binary square dot was used as a random dot pattern, they may be a dot pattern which attached a shade, for example, and the colored pattern. Although an example which may use it combining a mark of shape with various various sizes also about shape, and projects a pattern from the perpendicular upper part towards right under about geometric physical relationship of a pattern floodlight and an object body to an object was shown, Even if it projects from across, it is a thing which a depth map can detect similarly.

[0063] Although an example using a dot pattern which generated a position random as a pattern to floodlight was shown, When searching for corresponding points in block matching processing and it considers becoming almost horizontal search, even if it projects a random vertical slit group of a cycle, or a random vertical slit group of width as a pattern, for example, it cannot be overemphasized that same effect is done so.

[0064] Although a constitution method of a pattern floodlight was not explained in particular in detail, As conventional technology showed, for example, from it not being necessary to change a projection pattern while high degree of accuracy being unnecessary for projecting position accuracy of a random dot, and being processing, A device of easy unnecessary for example, composition like the usual slide show machine for home use is enough as a complicated projection device like a liquid crystal projector. This is one of the advantages on an equipment configuration of this invention.

[0065] Embodiment 2. explained operation as recognition objects supposing an object of a carton

box, the shape of i.e., a hand piece, as already stated, but. Even if it is an object of other shape, for example, a saccate object like a cement bag, composition of this invention can generate a depth map easily, and a three-dimensional position of a subject can be measured by a recognition procedure in said this example. About subsequent processings, it is the same as that of working example of the 1st invention as mentioned above than depth map generation. That is, final object position detection is possible by having an objective corner part and a straight-line contour part as a two-dimensional reference pattern, and unifying a positioning result with each recognition object. Thus, this invention can recognize objective three-dimensional information not only when objective shape is a polyhedron, but in the time of the shape and others like a bag.

[0066]A flow chart which shows drawing 8 and a flow of operation for a block diagram showing composition of embodiment 3. embodiment 3. is shown in drawing 9. Drawing 10 is an explanatory view of a situation where a level difference region part is extracted from a low resolution depth map, and only a level difference region part of a high resolution depth map carries out with stereo correspondences of high resolution. In drawing 8, 13-19 have the same function as embodiment 2., and explanation is omitted. A with means corresponding to a low resolution block by which 31 generates a depth map of a low resolution from a stereo image which the image input means 21 picturized with a low resolution, A level difference region extraction means by which 32 extracts a level difference region with a low resolution, a with means corresponding to a high resolution block by which 33 generates a depth map of high resolution about a level difference region part extracted from a picture of a low resolution, and 34 A low resolution depth map, It is a depth map synthesizing means which compounds a high resolution depth map of a level difference region part extracted from this depth map.

[0067]If a power supply is turned on and it starts by step ST301, by step ST302, a lamp power supply of the random texture pattern light projection means 20 will be turned on, and a random dot pattern will be projected to a recognition object thing. In step ST303, a stereo image of a picture of a couple is inputted into the with means 31 corresponding to a low resolution block, and the with means 32 corresponding to a high resolution block by the image input means 21 which comprises a camera of a couple. A stereo image turns off the random texture pattern light projection means 20 by step ST304, after an input finishes. In step ST305, a shade image of an object body is stored in a shade image storing means 15 from the first image input means 21a or second image input means 21b.

[0068]In step ST306, a stereo image is processed by the low resolution block matching means 31. Drawing 10 (a) is the figure which expressed the low resolution depth map portion A obtained by the with means 31 corresponding to a low resolution block in \*\* type. Although a CCD camera with a resolution of 512x512 pixels is used as an image input means, It is obtained because a 32x32-pixel depth map shall be generated as a low resolution depth map, and a method of block matching shall be performed in a method shown by embodiment 1., and a similar way, for example, block size shall be 16x16 pixels. In step ST307, the level difference region extraction means 32 extracts the level difference region portion B of a subject based on the above-mentioned low resolution depth map. A mimetic diagram of the level difference region portion B is shown in drawing 10 (b). In step ST308, a depth map of high resolution is generated more by the with means 33 corresponding to a high resolution block only about the above-mentioned level difference region portion (shadow area of B-A). In this case, high resolution is 128x128 pixels. While this processing sets block size as 8x8 in block matching processing and shifts 4 pixels of blocks at a time, it is obtained by detecting corresponding points, and the depth map C of a level difference region portion is shown in drawing 10 (c). In drawing 10 (c), a level difference region of an object in which a shadow area (within the limit [ of B ]) was obtained from a low resolution depth map, and a gray portion (within the limit [ of C ]) are portions equivalent to a level difference region portion of a depth map newly obtained as a result of high resolution depth map generation. In step ST309, the above-mentioned low resolution depth map and a high resolution depth map are compounded by the depth map synthesizing means 34, and a depth map of resolution of 128x128 like drawing 10 (d) is obtained.

[0069]It is processed like embodiment 2. after step ST310. By the highest rung surface area

extraction means 13, in Step 10, it is extracted from a depth map highest rung field equivalent to a highest rung object face, and in step ST311. A two-dimensional reference pattern is automatically generated by the two-dimensional reference pattern creating means 17 using size data of recognition objects stored in the object size database 16. In step ST312, a picture from which the above-mentioned highest rung surface area was extracted is divided into two, a highest rung side and the other field, and is binary-ized, and one of the highest rung object candidates is extracted by the highest rung object candidate extraction means 14. In step ST313, a precise position of candidate objects is measured and detected by the object position detecting means 18 [ near / which was extracted / one object candidate's position ] using the above-mentioned two-dimensional reference pattern and the above-mentioned shade original image information. Thus, a two-dimensional position of determined candidate objects is unified with depth map information on high resolution compounded [ above-mentioned ] by the information integration means 19, and is eventually outputted as objective three-dimensional information by it. In step ST314, by processing of step ST312 and ST313. If an object which will be that processing is ended by step ST315 and will not be recognized if it is judged that it carried out by making a judgment of whether all the objects have been recognized and no is judged to still exist, processing which returns to step ST312 and ST313 and extracts an object candidate will be repeated. What an objective position became final and conclusive one by one based on a binary-ized depth map from which the above-mentioned highest rung surface area was extracted is eliminated, and a judgment of whether all the objects have been recognized is performed by investigating whether highest rung surface area which moreover remains on a picture exists.

[0070]As mentioned above, in order that this embodiment 3. may obtain a depth map of high resolution eventually, A depth map of a low resolution is obtained beforehand for a short time, and it can decide to limit to a field equivalent to an objective step part, and to perform processing with stereo correspondences of high resolution, and can make it possible to generate a depth map of high resolution by little calculation time, and an object candidate's extraction precision etc. can be raised.

[0071]Although a low resolution depth map used 32x32 pixels and a high resolution depth map as 128x128 pixels, Different resolution from this may be set up in consideration of processing time permitted, a size of an object body to recognize, etc., and a merit by time of low resolution depth map generation of high speed processing and combination of processing with high resolution stereo correspondences is not lost.

[0072]Although it had composition by making a stereo image of a low resolution depth map generate time, and a stereo image of a high resolution depth map generate time into the same picture, Block matching may be performed based on a degeneration picture which acquired a picture of full resolution (the same resolution as CCD camera resolution) by degenerating one half, 1/4, etc. as a low resolution stereo image. A lens of a CCD camera of the first and the second image input means 21 is made to zoom in a high resolution depth map generate time, and it may be made to picturize more a step part of an object in a low resolution depth map in details.

[0073]Although a random dot pattern projected on an object body was used as a fixed pattern, A dot pattern floodlight of two kinds of different sizes is prepared, big dot pattern projection is performed to a low resolution depth map generate time, and if it changes and carries out so that smaller dot pattern projection may be performed to a high resolution depth map generate time, a high-precision depth map will be obtained.

[0074]A flow chart which shows drawing 11 and a flow of operation for a block diagram showing composition of embodiment 4. embodiment 4. is shown in drawing 12. In drawing 11, 13-21 have the same function as embodiment 2., and explanation is omitted. As for 41a, the first image storing means and 41b of a high-speed similarity calculation means and 43 are [ a corresponding block reliability judgment means and 45 ] depth map storing means a high reliability similarity calculation means and 44 the second image storing means and 42.

[0075]If a power supply is turned on by step ST401 and a device is started, a power supply of the random texture pattern light projection means 20 will be turned on by step ST402, and a



random dot pattern will be projected to a recognition object thing. In step ST403, a stereo image of an object body is picturized by the first image input means 21a and second image input means 21b, and it is inputted into the first image storing means 41a and second image storing means 41b by them, respectively. For a start, CCD of the resolution with two same cameras of the second image input means 21a and 21b is used, it is equipped with a lens of the same focal distance, and an optic axis is arranged at a predetermined distance. After an image input, a power supply of the random texture light projection means 20 turns off by step ST404, and a random dot pattern projected on an object disappears. In step ST405, a picture on which a random dot is not projected by the first image input means 41a is picturized, and it is stored in the shade original image storing means 15. This picture is used for object position detection later.

[0076] For a start, the second image input means 21a and 21b is usually arranged on the horizon in stereo \*\* at right and left, and is made to call a right camera the first image input means 21a, and to call a left camera the second image input means 21b. Suppose that a picture stored in a right image and the second image storing means 41b in a picture stored in the first image storing means 41a is called a left image in connection with it.

[0077] In step ST406, the high-speed similarity calculation means 42 performs block matching at high speed based on a stereo image stored in the first and the second image storing means 41a and 41b. In stereo vision, block matching considers it as a small region (block) which divided a unit with correspondence in the shape of a lattice, and carries out a picture on either side with correspondence for every block. It is as having indicated block matching to embodiment 1. Under the present circumstances, an accumulation difference absolute value is used as evaluation of similarity between two image blocks. For example, in a noticing block, when  $b_R(i, j)$  and a left image are set to  $L(i, j)$  and block size is made into  $N \times N$ , accumulation difference absolute value  $S_{SAD}(dx, dy)$  in a position  $(dx, dy)$  on a left image is expressed with (the formula 2).

[0078]

[Equation 2]

$$S_{SAD}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \quad \dots\dots\dots (式2)$$

[0079] To a certain  $b_R$ , it searches for  $L$  (left) picture and  $S_{SAD}$  detects the position  $(dx, dy)$  which becomes the minimum. In this way, the block  $b_L$  on the left image corresponding to the noticing block  $b_R$  is detected. If the pair of a corresponding block is discovered, the similarity of high reliability will be calculated by the above (formula 2) by the high-reliability similarity calculation means 43 by step ST407 to the pair. A left image is not scanned at this time. It receives that (the formula 2) could be evaluated  $(dx, dy)$ , and the following (formula 3) estimates similarity. Generally  $S_{CORR}$  is called a normalization cross correlation function.

[0080]

[Equation 3]



$$S_{\text{CORR}} = \frac{C}{A \cdot B}$$

$$A = \sqrt{N^2 \sum_{j=1}^N \sum_{i=1}^N \{L(d_x + i, d_y + j)\}^2 - \left\{ \sum_{j=1}^N \sum_{i=1}^N L(d_x + i, d_y + j) \right\}^2} \quad \cdots \text{ (式4)}$$

$$B = \sqrt{N^2 \sum_{j=1}^N \sum_{i=1}^N \{b_R(i, j)\}^2 - \left\{ \sum_{j=1}^N \sum_{i=1}^N b_R(i, j) \right\}^2} \quad \cdots \cdots \cdots \text{ (式5)}$$

$$C = N^2 \sum_{j=1}^N \sum_{i=1}^N \{L(d_x + i, d_y + j) \cdot b_R(i, j)\} \\ - \sum_{j=1}^N \sum_{i=1}^N L(d_x + i, d_y + j) \sum_{j=1}^N \sum_{i=1}^N b_R(i, j) \quad \cdots \cdots \text{ (式6)}$$

[0081](The formula 3) calculates similarity more exact compared with (the formula 2), and reliable. Obtained similarity  $S_{\text{CORR}}$  is stored in the depth map storing means 45 with the azimuth difference obtained from the with result corresponding to the above, i.e., distance information.

[0082]In step ST408, by the corresponding block reliability judgment means 44, the above — high reliability similarity is set up beforehand and it is compared with the similarity threshold stored in the inside of a computer, when similarity is large enough, it is judged with [ corresponding to the right ] having been made, and when similarity is smaller than a threshold, it is judged with it having been with the mistaken correspondence. A decision result is stored in the depth map storing means 45. When it carries out with the correspondence using similarity high-speed about all the blocks of a right image, and the similarity of high reliability and reliability with correspondence is not judged by step ST409 continuing, the processing more nearly same than processing step ST406 is repeated again.

[0083]It is processed like embodiment 2, after step ST410. By the highest rung surface area extraction means 13, in Step 410, it is extracted from a depth map by the highest rung field equivalent to a highest rung object face, and in step ST411. A two-dimensional reference pattern is automatically generated by the two-dimensional reference pattern creating means 17 using the size data of the recognition objects stored in the object size database 16. In step ST412, the picture from which the above-mentioned highest rung surface area was extracted is divided into two, a highest rung side and the other field, and is binary-ized, and one of the highest rung object candidates is extracted by the highest rung object candidate extraction means 14. In step ST413, the precise position of candidate objects is measured and detected by the object position detecting means 18 [ near / which was extracted / one object candidate's position ] using the above-mentioned two-dimensional reference pattern and the above-mentioned shade original image information. Thus, the two-dimensional position of the determined candidate objects is unified with the depth map information on the high resolution compounded [ above-mentioned ] by the information integration means 19, and is eventually outputted as objective three-dimensional information by it. In step ST414, by processing of step ST412 and ST413. If it is judged that the object which processing is ended by step ST415 and is not recognized still exists if it is judged that it carried out by making a judgment of whether all the objects have been recognized and no, the processing which returns to step ST412 and ST413 and extracts an object candidate will be repeated.

[0084]Above, like, by this embodiment 4., when carrying out a stereo image with correspondence by a block unit, By asking for the similarity of a corresponding block precisely by a means to

perform picture search using a high-speed similarity evaluation means, and to evaluate high reliability similarity only once first about a pair of a corresponding block which was able to be found, Search processing which time requires can be performed at high speed, and a corresponding block can perform a judgment of being the right by high reliability eventually.

[0085] Although normalization cross correlation was used as a similarity calculation equation of high reliability in this embodiment 4., using an accumulation difference absolute value as a high-speed similarity calculation equation, For example, the maximum of an absolute value of a difference of each pixel value during a block may be used as high-speed similarity, or total of a square of a difference may be used as high reliability similarity. Essence of embodiment 4. takes effect with combination of a similarity calculation means which attached greater importance than to reliability to rapidity, and a similarity calculation means with the reverse characteristic.

[0086] In high-speed similarity calculation and high reliability similarity calculation, although this embodiment 4. used each pixel value of an image block similarly, As a pixel used at the time of high-speed similarity calculation was called from a block 1/2 or 1/4 by infanticide operation, it may limit and choose, In that case, even if a similarity evaluation type used for the two above-mentioned kinds of similarity calculation even if is the same, since high-speed similarity calculation is performed at high speed, same effect is done so.

[0087] That this embodiment 4. performs all the processings by a computer A premise, When environment where similarity calculation of a certain kind is performed at high speed by available special hardware is provided, for high-speed similarity calculation, similarity may be calculated by a method by which another high-reliability is secured using the hardware at the time of high reliability similarity calculation.

[0088] A flow chart which shows drawing 13 and a flow of operation for a block diagram showing composition of embodiment 5. embodiment 5. is shown in drawing 14. In drawing 13, 13-20 have the same function as embodiment 2., and explanation is omitted. 51 is an image input means and the first image input means 51a, the second image input means 51b, and 51c the third three image input means have it. 52 is a with means corresponding to a stereo image block.

[0089] When a power supply is turned on by step ST501 and a device starts, a random texture pattern light projection means is turned on by step ST502, In Steps ST503-ST505 on which a random dot pattern is projected to a recognition object thing, a picture of three sheets is inputted one by one by the first image input means 51a, the second image input means 51b, and the third image input means 51c. What has the same optical specification which the three image input means 51a, 51b, and 51c are CCD cameras, and these three cameras are the same resolution, and was equipped with a lens of the same focal distance is used. Arrangement of a camera is arranged in parallel so that an optic axis of each camera may be on the same flat surface on space and no optic axis may cross. Here, three cameras will be called for convenience a left camera, a center camera, and a right camera, respectively. Distance between optic axes arranges an interval of a left camera and a center camera equally to an interval of a center camera and a right camera. This camera arrangement is equivalent to having arranged the 3rd camera in the center of two stereo arrangement cameras in above-mentioned embodiment 2.

- 4.  
[0090] A picture is stored in an image storing means which is a different image storing means, respectively, a lamp power supply of the random texture pattern floodlight 20 turns it off by step ST507, and a random dot pattern of an object surface disappears at the same time an image input is completed. In Step 407, a shade image of a subject is picturized by the second central image input means 51b, and it is stored in the shade original image storing means 15 by it.

[0091] In step ST508, the with means 52 corresponding to a stereo image block performs processing with correspondence of a stereo image. About flowing into this processing, a flow chart of detailed operation is shown in drawing 15. Stereo-correspondences attachment processing starts by ST508-1, and step ST508-2 of the beginning is searched for corresponding points between a picture by the first image input means 51a (left camera), and a picture by the second image input means 51b (center camera). It is with processing corresponding to a block of two pictures, and this processing is performed like processing shown by the above-mentioned Embodiment 2. An image pick of a center camera is divided into a block small region, and this

embodiment 5. is searched for corresponding points on a left camera for every block. Drawing 16 shows a situation of a with corresponding to a block between these two pictures in \*\* type. Drawing 16 (a) shows a picture in which a carton box as recognition objects is reflected. In the case of a packed-up carton box, a thing on a transparent film is often stuck on the surface. In that case, a block judged like drawing 16 (b) to be information lack may occur by light from the pattern floodlight 20 or other indoor lighting systems, especially irregular reflection light. With this information missing part, for example in the center camera 51b, there is the state where it has happened in the left camera 51a although scattered reflection has not happened to a tape part, and a fact that an image pattern corresponding to the block concerned cannot discover in a picture of another side occurs. Processing in this state is performed by judging whether it is an information lack block in step ST508-3.

[0092]In step ST508-4, with [ corresponding to a block ] is performed between the second picture that a center camera picturized, and a right camera or the third picturized picture. A picture by a center camera is divided into a block, and it searches for a point of corresponding for every block, on a right camera picture. An information lack block occurs like the above-mentioned corresponding point searching also at this time. In step ST508-5, information missing data of the above-mentioned left camera 51a and the center camera 51b is together put for information about a point [ missing / information ] acquired in combination of the above-mentioned center camera 51b and the right side camera 51c. At this time, there is no direct relation between a position of an information lack block in a left-central pair and a position of an information lack block in a central-right pair in any way. Also in which pair, since a left or right camera picture top is searched on the basis of a picture of the center camera 51b, information lack block information of each pair is given by a block unit on a center camera picture, and can be compared easily.

[0093]If it judges that a block judged as information being missing in a left-central pair now is not information lack in a central-right pair, it will suppose that it is not the block concerned information lack, and will be calculated using a pair which is not information lack as the parallax information. When both pairs are judged to be information lack, the block concerned is judged as information being missing, and parallax information is not stored. Same processing is performed to a whole block of a center camera, and stereo-correspondences attachment processing is completed by step ST508-6. That is, unless it is an information lack block in both a left-central pair and a central-right pair, it is possible to acquire parallax information, i.e., distance information.

[0094]As a similarity calculation means used in points-of-comparison search in stereo-correspondences search, accumulation difference absolute value  $S_{SAD}(dx, dy)$  of the following (formula 7) is used.

[0095]

[Equation 4]

$$S_{SAD}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \quad \cdots \cdots \cdots \text{(式 7)}$$

[0096]Next, in step ST509 of drawing 14, the highest rung surface area extraction means 13 extracts the field of an objective highest rung side using the depth map generated in the with processing corresponding to the above-mentioned stereo block.

[0097]It is processed like embodiment 2. after step ST510. In step ST510, a two-dimensional reference pattern is automatically generated by the two-dimensional reference pattern creating means 17 using the size data of the recognition objects stored in the object size database 16. In step ST511, the picture from which the above-mentioned highest rung surface area was extracted is divided into two, a highest rung side and the other field, and is binary-ized, and one of the highest rung object candidates is extracted by the highest rung object candidate extraction means 14. In step ST512, the precise position of candidate objects is measured and

detected by the object position detecting means 18 [ near / which was extracted / one object candidate's position ] using the above-mentioned two-dimensional reference pattern and the above-mentioned shade original image information. Thus, the two-dimensional position of the determined candidate objects is unified with the depth map information on the high resolution compounded [ above-mentioned ] by the information integration means 19, and is eventually outputted as objective three-dimensional information by it. In step ST513, by processing of step ST511 and ST512. If the object which will be that processing is ended by step ST514 and will not be recognized if it is judged that it carried out by making a judgment of whether all the objects have been recognized and no is judged to still exist, the processing which returns to step ST511 and ST512 and extracts an object candidate will be repeated.

[0098]As explained above, in this embodiment 5. 2 sets of stereo images are generated using three cameras of the image input means 51, to an information lack block produced with a stereo image of one group, by parallax information by a stereo image picture of other groups, information is complemented and lack of information is prevented.

[0099]In this embodiment 5., have arranged so that it may have an optic axis which exists in the same flat surface in parallel with regular intervals as spatial arrangement of three image input means, but. Even if arrange a camera optical axis of the left, the right, or its both so that a center camera optic axis may be crossed, or a camera interval is arranged so that it may not be regular intervals, or it arranges so that three optic axes may not necessarily be found on the same flat surface, an effect which a lack block complements similarly is acquired.

[0100]Although normalization cross correlation was used as a similar formula of high reliability in embodiment 5., using an accumulation difference absolute value as a high-speed similarity calculation equation, For example, the maximum of an absolute value of a difference of each pixel value during a block may be used as high-speed similarity, or total of a square of a difference may be used as high reliability similarity.

[0101]A flow chart which shows drawing 17 and a flow of operation for a block diagram showing composition of embodiment 6. embodiment 6. is shown in drawing 18. In drawing 17, 11-13, 15-17, 19, and 21 have the same function as embodiment 1., and explanation is omitted. As for 64, in a figure, a hypothetical validation means and 69 are object position detecting means an object candidate combination hypothesis generating means and 65.

[0102]By step ST601, when a device is started, with two cameras of the image input means 11, a picture of an object body is inputted into the depth map creating means 12, and is stored individually. A thing of the same specification between which two cameras do not share an optic axis is used. By step ST602, stereo matching of the stereo image of the two above-mentioned sheets is carried out by the depth map creating means 12, and a depth map is generated from the azimuth difference.

[0103]By step ST603, a shade original image picturized by the image input means 11 is stored in a shade original image storing means. A shade original image of an object made into a recognition object of this embodiment 6. is shown in drawing 19 (a). An object is a box-like object of a carton box etc., and they are loaded into several steps. A depth map is extracted from a picture of an upper surface portion of a highest rung object by \*\* et al. and the highest rung surface area extraction means 13 as a field by step ST604. Drawing 19 (b) is the extracted highest rung object area. Fields other than the highest rung are also shown like a figure.

[0104]In step ST605, a reference pattern which expressed an objective two-dimensional image pattern by the two-dimensional reference pattern creating means 17 is automatically generated using data stored in the object size database 16 which stored size data of a case shape object. A mimetic diagram of the reference pattern is shown in drawing 19 (c).

[0105]Explanation of the following operations is a portion used as essence of this embodiment 6. In step ST607, one combination hypothesis of an object candidate is generated from highest rung surface area and a two-dimensional reference pattern which were extracted [ above-mentioned ] by the thing candidate combination hypothesis generating means 64. As a hypothesis was shown in drawing 19 (c), it calls this hypothesis "the hypothesis 1" in this embodiment 6. As opposed to what expressed a highest rung object area [ like drawing 19 (b) ] whose hypothesis is by a picture, A reference pattern of a recognition object thing is processed

by technique of pattern matching, for example, template matching, it is obtained by guessing combination of a two-dimensional reference pattern which forms highest rung surface area like drawing 19 (b). An object two-dimensional position is simultaneously detected like drawing 19 (d) by the object position detecting means 69. Comparison with a shade original image and the above-mentioned two-dimensional reference pattern which were stored [ above-mentioned ] performs object position detection. It is template-matching \*\*\*\* between edge images detected from the above-mentioned shade image using information on a contour part of a two-dimensional reference pattern. In a pattern searching field in template matching processing since an objective coarse position is known in a stage of hypothetical generation in that case, only the above-mentioned neighborhood of a position is.

[0106]It is verified in step ST608 whether the above-mentioned hypothesis is appropriate. It is calculated whether the hypothetical validation means 65 generates a hypothetical picture like drawing 19 (d), and highest rung surface area like drawing 19 (b) may be rationally generated by combination of six objects to the object candidates 1-6. That is, a hypothetical picture generated by the six above-mentioned object candidates is compared with a highest rung side image, and it operates so that it may judge with it being the hypothesis which computed area of a field used as the difference, and was mistaken, so that the above-mentioned area was large. In step ST608, the one validity index is given to one hypothesis generated by step ST606 in this way. A reciprocal of the above-mentioned area serves as a validity index.

[0107]By step ST609, repeatedly, by step ST610, all hypotheses are generated and a hypothesis with the highest validity is chosen in them until it generates all hypotheses that can consider processings from step ST606 to ST608. That is, said validity index about each hypothesis is compared, and a hypothesis with the highest index is chosen. Drawing 19 (e) and (f) is the respectively different hypothesis 2 and the hypothesis 3. Since the hypothesis 2 has the highest validity index in this case so that clearly from a figure, this will be chosen in step ST610 as best hypothesis.

[0108]In step ST611, two-dimensional position information equivalent to the above-mentioned best hypothesis and said depth map are unified by the information integration means 19, it is outputted from a device as a recognition result by step ST612, and processing is ended by it.

[0109]Although stereo \*\* with two cameras was used for this embodiment 6. as a depth map creating means, A depth map creating means by stereo \*\* which used together floodlighting of characteristic floodlighting patterns, such as stereo \*\* with three or more cameras and a random dot, or a spatial code-sized method explained as conventional technology may be sufficient. Since it has the feature that this Embodiment 6 verifies a hypothesis of a recognition result obtained from a depth map, and obtains a result which will seemingly be reasonable, what kind of thing may be sufficient as a generation method of a depth map?

[0110]Although a method of calculating a validity index based on a difference of a hypothetical picture and a actual picture was used for this embodiment 6. as a method of verifying hypothetical validity, For example, same effect can be acquired, even if it uses an index which is made high [ validity ] so that total to all the candidate objects of similarity of template matching calculated when acquiring the above-mentioned hypothesis is large.

[0111]Although profiling information of an object obtained from a reference pattern and an edge image generated from a shade original image were processed with the template-matching technique as a detecting method of an object position in this embodiment 6., Same effect can be acquired even if it uses the positioning techniques, such as template matching of a binary format image, for this.

[0112]Although this embodiment 6. explained a process flow which performs an object candidate's generation and verification of a combination hypothesis one by one, many hypotheses are generated at once and stored, and those validation is performed at once and it may be made to choose what has the highest validity.

[0113]Although embodiment 6. explained that all the hypotheses were generated, since a hypothesis is acquired as an object candidate's combination, it is treated as an objective combinatorial optimization problem, and same effect is acquired even if it calculates an approximate solution.

[0114] A flow chart which shows drawing 20 and a flow of operation for a block diagram showing composition of embodiment 7. embodiment 7. is shown in drawing 21. In drawing 20, 11-13, 15-19 have the same function as embodiment 6., and explanation is omitted. As for 71, an object arrangement judging means and 73 are warning generation means an object arrangement reference database and 72.

[0115] This embodiment 7. is constituted so that warning may be generated, when object arrangement is accidentally arranged by composition of embodiment 6. Therefore, Steps ST701-ST711 of a flow of drawing 21 are the same as those [ 601-611 ] of drawing 18.

[0116] When a device is started by step ST701, by the image input means 11, a picture of an object body is inputted into a depth map creating means, and is stored individually. By step ST702, stereo matching of the picture of the two above-mentioned sheets is carried out by the depth map creating means 12, and a depth map is generated from the azimuth difference. It may be the stereo vision which used together a pattern floodlight as the necessity that this processing is not necessarily simple stereo matching not had, for example, shown in embodiment 2.

[0117] In step ST703, a shade original image of a recognition object thing is inputted by the image input means 11, and it is stored in the shade original image storing means 15. Highest rung surface area is extracted by the highest rung surface area extraction hand 13 by step ST704. In step ST705, a reference pattern which expressed an objective two-dimensional image pattern by the two-dimensional reference pattern creating means 17 using the object size database 16 which stored size data of a case shape object is generated automatically. In step ST706, one combination hypothesis of an object candidate is generated by the object candidate combination hypothesis generating means 64 from highest rung surface area and a reference pattern which were extracted [ above-mentioned ]. As [ showed / a hypothesis / that embodiment 6. explained / same drawing 19 (d) ] As opposed to what expressed a highest rung object area [ like drawing 19 (b) ] whose hypothesis is by a picture in this embodiment 7., A reference pattern of a recognition object thing was processed by technique of pattern matching, for example, template matching, and it has obtained by guessing combination of a two-dimensional reference pattern which forms highest rung surface area like drawing 19 (b). In step ST707, a two-dimensional position of each candidate objects contained in the above-mentioned hypothesis by the object position detecting means 69 like drawing 19 (d) is detected. Object position detection is detected by template matching between a shade original image stored [ above-mentioned ] and an edge image detected from a shade image using information on a contour part of the above-mentioned two-dimensional reference pattern.

[0118] In step ST708, a hypothetical picture generated by two or more above-mentioned object candidates is compared with a highest rung side image, and it judges with a possibility of being the hypothesis which computed area of a field used as the difference, and was mistaken, so that it was large being high. In Step 709, repeatedly, at Step 710, all hypotheses are generated and a hypothesis with the highest validity is chosen in them until it generates all hypotheses that can consider processing of step STs 706-708. That is, said validity index about each hypothesis is compared, and a hypothesis with the highest index is chosen.

[0119] In step ST711, two-dimensional position information and the above-mentioned depth map which are equivalent to the best hypothesis by the above-mentioned hypothesis by the information integration means 19 are unified. At this time, it means that an image processing device in this embodiment 7. had known objective arrangement and a three-dimensional position of each object, and it is stored in an inside of a device. In Step 712, it is judged whether a recognition result stored [ above-mentioned ] has agreed compared with right array data prepared beforehand by the object arrangement judging means 72. Right array data is the arrangement pattern decided with an object stored in the object arrangement reference database 71 to be recognized. for example, -- in a physical distribution factory where this device to embodiment 7. twist is used -- a carton box as recognition objects -- that shape and size -- beforehand -- the right -- " -- how to stack a box to a palette top which stacks and is called pattern" is decided on.

[0120] Objective arrangement which should be made such a standard is described by the above-



mentioned object arrangement reference database. When [ from which a result recognized in step ST712 serves as a standard ] stacking and not having agreed with a pattern, in step ST713, it tells that generate warning to a device user or an administrator of a host system, and an object is not correctly stacked by the warning generation means 73. Warning is \*\*\*\* made a sound according [ this embodiment 7. ] to a buzzer by character representation to a display device for a display connected to a device. When judged with a recognition result being a right pattern in step ST712, gripping operation of an object by a robot will be performed by a final object recognition result's being outputted by the information integration means 19, for example, transmitting data to a robot. As recognition operation, operation is ended by step ST715.

[0121]In this embodiment 7., when it not only recognizes an object in this way, but it judges whether it is the arrangement expected by comparing that arrangement pattern with criterion data and the above-mentioned recognition result differs from criterion data further, it has the function to generate warning. There is an advantage that he can stack, and a worker can notice a work error by this, or intense collapse of cargo piles under transportation can be detected.

[0122]A method of verifying stereo \*\* with two cameras based on a difference of a hypothetical picture and a actual picture as the method of hypothetical validation as a depth map creating means and the template-matching technique which used objective profiling information as a detecting method of an object position are used, respectively. However, even if it realizes these partial processing means processing means with other techniques, it cannot be overemphasized that same effect is done so. For example, stereo \*\* which uses together a pattern explained by embodiment 2. as a depth map creating means, The template-matching method by a shade image pattern and normalization cross correlation which prepared beforehand each object contained in a hypothesis by making into an object position detecting method a method which is judged that validity is high, so that total of similarity [ a shade reference pattern / object / each ] is high as the hypothetical validation method, It may adopt, respectively.

[0123]When it judged that a recognition result does not suit as compared with an object arrangement reference database stored beforehand, an example which generates warning was explained, but the contents of warning are changed and it may be made to make a device user and a worker know according to a similar degree with criterion data. For example, it displays "Rightness of a way which was judged as a result of automatic recognition and to stack is the level 8" etc. on a control display for indication. As a result of having recognized furthermore, when a load collapse state of a cargo is judged that are dramatically remarkable and it is not appropriate to continue work as it is, it not only may emit the above-mentioned warning, but it may require a stop of a system of an administrative computer of a higher rank, or this device may be made to carry out a direct system stop.

[0124]Although a character representation by buzzer and a display for indication was adopted as a warning means, a method of generating human being's compounded voice may be used as a means to send warning to human being.

[0125]A flow chart which shows drawing 22 and a flow of operation for a block diagram showing composition of embodiment 8. embodiment 8. is shown in drawing 23. In drawing 22, 11-15, 17, and 19 have the same function as embodiment 1., and explanation is omitted. An artificial reference pattern creating means in which 81 generates an artificial two-dimensional reference pattern based on objective size data in a figure, An actual image reference pattern storing means and 88 are object position detecting means an actual image reference pattern logging means by which 82 extracts a reference pattern of an actual image corresponding to a size of recognition objects from a shade original image, and 83.

[0126]If a device is started by step ST801, a picture of recognition objects will be inputted into the depth map creating means 12 by the image input means 11. In step ST802, a depth map is generated by the depth map creating means 12 using an inputted picture. Although what kind of thing may be used as a depth map creating means, suppose that stereo vision with two cameras is used in this embodiment 8. Therefore, an image input means uses two television cameras.

These two cameras have the same optical specification, and they are arranged so that an optic axis may not be shared. Usually, it arranges so that a camera optical axis may become almost parallel like human being's eyes. In step ST803, one camera of the described image input means



11 is used, a shade original image is picturized, and it stores in the shade original image storing means 15. Even if a shade image picturized to the above-mentioned depth map generate time is used for this original image as it is, it does not interfere.

[0127]In step ST804, the highest rung surface area extraction means 13 extracts objective highest rung surface area using a depth map. A field of an extracted highest rung object is the same as that of what is drawn on drawing 3 (b) in \*\* type. That is, a highest rung field is expressed as a picture of a binary "0" So that "1" and the other field may be said. In this embodiment 8., it supposes that it is a recognition object a box-like object like a carton box, and the same object piled on the ground in heaps is picturized by an image input means installed downward in a ceiling, i.e., a television camera. In step ST805, an artificial two-dimensional reference pattern is generated by the artificial reference pattern creating means 81 based on size data of a box-like object used as a recognition object beforehand stored in the object size database 17. Although this is the same as that of a two-dimensional reference pattern in explanation of embodiment 1., in order that this may emphasize that it is the pattern generated artificially and automatically, especially by this embodiment 8., it calls in this way. This reference pattern expresses only a contour part of an object as shown in drawing 3 (b).

[0128]In step ST806, one candidate of an object considered to exist in the highest rung by the highest rung object candidate extraction means 14 is extracted from a binary format image expressing extracted highest rung surface area. Here, a candidate is extracted by detecting a peripheral area at which it crosses right-angled from an object being visible to a rectangular pattern on a picture to a highest rung object area expressed with a binary. A candidate's extracted mimetic diagram is the same as that of drawing 3 (d) in explanation of embodiment 1. In step ST807, the above-mentioned object candidate is positioned by using the template-matching method by the object position detecting means 14. The above-mentioned artificial reference pattern is used as a template. In template matching, an edge image obtained as a result of carrying out the edge detection process of the original image stored in the above-mentioned shade original image storing means 15 into the object position detecting means 88 as a picture used as an object which looks for a template is used. In this template matching, since a rough position is already detected in a stage of the above-mentioned candidate extraction, just processing in which only that very neighborhood is searched from a candidate detection position is enough as matching here. A situation of positioning is as drawing 3.

[0129]As a result of positioning processing, although similarity with a template is obtained with position information, this is data expressing similarity to an artificial reference pattern generated beforehand. An object detected, so that similarity was high resembles a right box assumed beforehand, and when similarity is low, it can be said that a possibility that positioning has gone wrong is high. So, in step ST808, it \*\*\*\* that it is the reliability of object detection of obtained similarity data, and this judges whether it is larger than a reliability threshold beforehand stored in this device. Reliability of an object obtained as a result of positioning judges that an object mistaken when smaller than a reliability threshold was detected, and repeats processing from step ST806, another highest rung object candidate is extracted again, and a detecting position using an artificial pattern is performed. When obtained reliability is larger than a threshold, it will progress to the following step ST809.

[0130]In step ST809, by the actual image reference pattern logging means 82, a reference pattern of an actual image corresponding to a size of recognition objects in an object position detected eventually is extracted, and it is stored actual image reference pattern storing means 83. At this time, an actual image means a shade original image stored in the shade original image storing means 15, and a logging position is a position of an object detected [ above-mentioned ]. A logging field is a rectangle and the size is changed into a size on a picture with distance of a camera and an object which were able to obtain dressed-size data of the above-mentioned object from the above-mentioned depth map. That is, a rectangular image field as a shade image corresponding to one object to recognize is started.

[0131]In step ST810, an object after the 2nd is detected from a picture by the object position detecting means 88. Here, different techniques from the time of said first object detection are used. A pattern in which even an object as a shade actual image stored in the above-mentioned

actual image reference pattern storing means 83 corresponds is made to a template, and the template-matching method for detecting a pattern position which resembles a template most out of an original image stored in the shade original image storing means 83 is applied.

[0132]It is judged whether it is equivalent to all the objects required as liking to recognize a detected object beforehand in step ST811. If all the objects are recognized, after calculating a three-dimensional position of each object by unifying the above-mentioned detection object position and depth map information and transmitting the result to an upper rank host computer or a robot by the information integration means 19, processing is ended by step ST812. Processing is repeated from step ST810 until objective recognition finishes. That is, although an object of eye one is recognized using a depth map etc., an object after the 2nd is detected by template matching based on an actual image pattern as a shade image obtained at the time of the 1st recognition.

[0133]Although an example which adopted general stereo vision as a depth map creating means explained in embodiment 8., An effect that depth map generation by stereo \*\* using random dot pattern floodlighting which was explained by embodiment 2., and a spatial code-ized method explained by a Prior art is also the same is acquired.

[0134]A flow chart which shows drawing 24 and a flow of operation for a block diagram showing composition of embodiment 9. embodiment 9. is shown in drawing 25. in drawing 24 -- 11 -- an image input means and 91 -- as for a depth map creating means and 92a, the first depth map storing means and 92b of a depth map comparison means and 94 are [ a position-correction-quantity calculating means and 96 ] object transportation devices an initial gripping position calculating means and 95 the second depth map storing means and 93.

[0135]Objective shape is a polyhedron like a carton box so that explanation may be easy to be understood, and it is explained supposing a thing of identical shape and an identical size. This embodiment 9. can use composition of this embodiment 9., even if shape of recognition objects does not necessarily need to be polyhedral form and objects are saccate objects, such as a cement bag.

[0136]If a device is started by step ST901, a picture of an object of a recognition object will be inputted into the depth map creating means 91 by the image input means 11. In step ST902, using an inputted picture, a depth map is generated by the depth map creating means 91, the first depth map is generated in step ST903, and it is stored in the first depth map storing means 92a. Although what kind of thing may be used as a creating means of a depth map, there are a spatial code-ized method and a method of generating from stereo vision, for example.

[0137]Drawing 26 explains a depth map. An original image picturized by the image input means 11 in the state where a box-like object when initial-state, i.e., step ST, 901 was performed was tiered is expressed by drawing 26 (a) in \*\* type. Drawing 26 (b) expresses a depth map stored in the first depth map storing means 92a in \*\* type. A depth map expresses distance from a camera of an image input means to an object in picture, expresses it with such a bright color that distance from a camera is near, and is expressed in such a dark color tone that distance from a camera becomes large.

[0138]In the following step ST904, an initial gripping position is calculated by the initial gripping position calculating means 94 from a depth map. An initial gripping position is a position grasped in order to move an object by the object transportation devices 96, such as a robot. Let a position which processed the first depth map of the above and has recognized a corner part from a field currently displayed in a bright color in an object area (b), i.e., drawing 26, which are located in the highest rung and to which fixed offset was applied from a corner position for determining an initial gripping position be an initial gripping position. In this case, it means that it was set up on the object A. As for the direction of an arrow, an intersection of a grip direction, an arrow, and a straight line expresses a gripping position on drawing 26 (b). However, even if a gripping position which did not necessarily determine a gripping position that the object A should be grasped intentionally here and for which it asked by a simple method from a depth map exists on objects other than the object A in this case, If an initial gripping position was set up by something on a same thing object as a result, in transfer operation of a cargo, it will not be a problem at all.

[0139]In Step 905, an object change means slack robot grasps an initial gripping position, and an object is temporarily evacuated out of a view of an image input means. Drawing 26 (c) shows a situation after object evacuation. A robot adsorbs a gripping position on the object A by an adsorption hand, and objective saving operation is performed by moving a hand to a retreating position decided beforehand. However, a robot is evacuated by making a hand stand it still, with the object A grasped here. A hand is the position which understands that a position which is making the object stand it still is out of a view of the described image input means 11 clearly, and a retreating position decided beforehand is a position set up such.

[0140]A depth map of a cargo is again generated by the depth map creating means 91 after object evacuation, and it stores in the second depth map storing means 92b at Step 907 step ST906. Drawing 26 (d) expresses the second depth map of this embodiment 9. in \*\* type. In this figure, a field which is equivalent to the evacuated object A from a highest rung object area has applied.

[0141]In step ST908, the first depth map and second depth map are compared by the depth map comparison means 93, and position correction quantity is further calculated by the position-correction-quantity calculating means 95 based on the result. In this embodiment 9., difference simple as comparison of a depth map is performed. Drawing 26 (e) is detected from the first depth map by lengthening the second depth map. The field B in which this figure is equivalent to the object A evacuated temporarily [ drawing 26 (c) ] remains as a difference of two depth maps. The number of objects grasped and evacuated by a robot is one, and, naturally only a portion which is equivalent to an evacuated object in a depth map before and after saving operation exists as a difference. Since it turns out that the field B is equivalent to an object piece as a difference result of a depth map, the shape and centroid position are easily calculable. After binary-izing a difference image clearly on a picture, for example as a background and the technique of calculating the center of gravity of a field, shape, and direction which were separated and performing labeling for every field, a technique generally used from the former by which each label attachment was carried out, and which calculates the center of gravity and moment maximum directions for every field is applicable.

[0142]Although drawing 26 (f) is an enlarged drawing of an evacuated object, a gripping position which was obtained from a depth map difference result and which was calculated newly can be found besides an initial gripping position, and it is taken as a gripping position which calculated the above-mentioned centroid position newly. Position correction quantity is calculated as a difference of these two gripping positions.

[0143]Although a robot has still grasped an initial gripping position at this time, a difference of a right gripping position, i.e., an objective centroid position, is already calculated. Therefore, a hand will know which objective position is grasped by sending the above-mentioned position-correction-quantity data to a robot about the object A which evacuated a robot in step ST909. Therefore, in the case of moving operation by a continuing robot, the robot can perform moving operation in consideration of the above-mentioned correction amount. This is equivalent to having already known correctly for a robot, in moving work of a cargo, when an object was grasped first, objective centroid position, i.e., position which should be grasped, It means having realized operation same with a robot grasping an objective exact gripping position and performing moving work.

[0144]Supposing it judges whether all the objects have been recognized and recognizes by step ST910, it will progress to step ST911 and processing will be finished. When processing has not been ended, processing is repeated from step ST902 until processing is completed.

[0145]As mentioned above, gestalt 9. evacuates to this operation based on an initial gripping position recognized from the first depth map, and evacuates only one object out of a picture, When it becomes a difference with the second depth map generated after that, an exact gripping position of an object under present evacuation is calculated based on shape and position information of an appearing object and a robot moves an object to the target position, a difference of an initial gripping position and an exact gripping position is amended.

[0146]Although it was considered as a position to which fixed offset was applied from corner points and a gripping position was determined as the technique of asking for an initial gripping

position in this embodiment 9. based on a corner part of the first depth map to a highest rung object area. For example, it may consider that a highest rung field in a depth map and the other field are binary format images, and a gripping position may be determined using the binary template-matching method.

[0147] Although operation to which it is made to move grasping an initial gripping position by a robot hand, and grasping an object by a hand out of a picture view as objective temporary evacuation was performed in this embodiment 9., an object may once be taken down to a predetermined place. If an initial gripping position can be grasped when grasping an object by a hand again after taking down, the same operation as this embodiment 9. can be considered.

[0148] Although it was considered as operation which makes a robot hand which grasped an object stand it still by a position as saving operation in this embodiment 9., By performing processings from step ST906 in drawing 25 to ST908 at high speed, it can also serve as saving operation with operation which moves a hand continuously. That is, it takes out out of an objective picture view, while moving a robot hand near [ purpose slack transfer place ] moving work, position-correction-quantity calculation is completed, and when a robot sets an object at a predetermined place for a transfer, even if it performs operation which considered position correction quantity, it has the same effect.

[0149]

[Effect of the Invention] The depth map creating means in which the image processing device concerning Claim 1 of this invention generates the depth map of two or more objects, A highest rung surface area extraction means to extract objective highest rung surface area from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, The two-dimensional reference pattern creating means which generates the two-dimensional reference pattern on an objective two-dimensional picture based on the size data of recognition objects, The shade original image storing means which stores the shade original image inputted from the camera, The object position detecting means which detects an objective position using the information on a two-dimensional reference pattern and a shade original image about the highest rung object candidate extracted individually, By having had the information integration means which unifies the detection result of the object position obtained by each means, and the distance information of each object obtained by the depth map creating means, It may be small as a device, since the image input means of recognition objects and a depth map creating means are the things of coarse accuracy and end, and cheap, and a floodlighting device may also be simple and its reliability improves.

[0150] The depth map creating means in which the object transfer equipment concerning Claim 2 of this invention generates the depth map of two or more objects, A highest rung surface area extraction means to extract objective highest rung surface area from a depth map, A highest rung object candidate extraction means to dissociate separately and to extract an object from highest rung surface area, The two-dimensional reference pattern creating means which generates the two-dimensional reference pattern on an objective two-dimensional picture based on the size data of recognition objects, The shade original image storing means which stores the shade original image inputted from the camera, The object position detecting means which detects an objective position using the information on a two-dimensional reference pattern and a shade original image about the highest rung object candidate extracted individually, By having combined with the object transfer equipment the image processing device provided with the information integration means which unifies the detection result of the object position obtained by each means, and the distance information of each object obtained by the depth map creating means, and having carried out it in this way, An object transfer equipment it may be small as a device, since the image input means of recognition objects and a depth map creating means are the things of coarse accuracy and end as an object transfer equipment, and cheap, simple a floodlighting device and reliable is obtained.

[0151] The random texture light projection means which floodlights a texture pattern random [ image processing device / concerning Claim 3 of this invention ] in a depth map creating means, Since it constituted from a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the first image input means and the

second image input means which input a stereo image, and the first and the second image input means picturized, Since a random dot pattern is floodlighted to recognition objects, even if it has stuck the object of the solid color which does not have a pattern in the surface, or the label which light reflects in the surface, The distance data in that of a surface Gentleman value can be obtained correctly, the composition of a floodlight is also simple, it is small and cheap, and reliability becomes high well.

[0152]The random texture light projection means which floodlights a texture pattern random [ transfer equipment / concerning Claim 4 of this invention / object ] in a depth map creating means, The first image input means and second image input means that input a stereo image, By having combined with the object transfer means the image processing device which comprised a with means corresponding to a stereo image block to carry out with correspondence between two pictures which the first and the second image input means picturized, and having carried out it in this way, Since a random dot pattern is floodlighted to recognition objects, even if it has stuck the object of the solid color which does not have a pattern in the surface, or the label which light reflects in the surface, Surface each position is acquired correctly, the objective detection time as an object transfer equipment is short, an objective position is detected correctly, a transfer of an object can be performed correctly quickly, the composition of a floodlight will also become small and cheap and its reliability will improve.

[0153]The low resolution depth map creating means in which, as for the image processing device concerning Claim 5 of this invention, a depth map creating means generates a depth map with coarse resolution, A level difference region extraction means to extract an objective level difference region part from a low resolution depth map, Since it should comprise a high resolution depth map creating means which generates the depth map of high resolution to a level difference region, and a depth map synthesizing means which compounds two depth maps from which the resolution of a low resolution depth map and a high resolution depth map differs, There is an effect which the depth map of high resolution is generated and can also make an object candidate's extraction precision high by short calculation time by few image memories.

[0154]The low resolution depth map creating means in which, as for the object transfer equipment concerning Claim 6 of this invention, a depth map creating means generates a depth map with coarse resolution, A level difference region extraction means to extract an objective level difference region part from a low resolution depth map, The high resolution depth map creating means which generates the depth map of high resolution to a level difference region, Combine with an object transfer means the image processing device which comprised a depth map synthesizing means which compounds two depth maps from which the resolution of a low resolution depth map and a high resolution depth map differs, and by short calculation time. The depth map of high resolution is obtained, an object candidate's extraction time is short, extraction precision is high, an objective position is detected correctly, and a transfer of an object can be performed correctly quickly.

[0155]A high-speed similarity calculation means by which, as for the image processing device concerning Claim 7 of this invention, the with means corresponding to a block searches for a correspondence image block on either side at high speed, It is what comprised a high reliability similarity calculation means to calculate similarity to high reliability about the pair of the block which is search results, and a corresponding block reliability judgment means will judge that have no corresponding block if high reliability similarity is below a fixed threshold, Since the depth map was generated, a depth map highly precise at high speed is obtained, the reliability of picture matching becomes high, and the reliability of the whole device becomes high.

[0156]A high-speed similarity calculation means by which, as for the object transfer equipment concerning Claim 8 of this invention, the with means corresponding to a block searches for a correspondence image block on either side at high speed, A high reliability similarity calculation means to calculate similarity to high reliability about the pair of the block which is search results, The image processing device which comprised a corresponding block reliability judgment means judge that has no corresponding block when high reliability similarity was below the fixed threshold, Since it combines with an object transfer means and the depth map highly precise at high speed was obtained, an object candidate's extraction time is short, extraction precision is

high, the reliability of picture matching becomes high, an objective position is detected correctly, and a transfer of an object can be performed correctly quickly.

[0157]The image processing device concerning Claim 9 of this invention has three or more image input means, The with means corresponding to a block chooses two or more kinds of picture pairs which chose the inner 2 \*\* picture of each picture acquired from three or more image input means, Since the combination of a stereo image is changed and it is complemented even if the lack block of the pair of a stereo image arises, since with [ corresponding to a block ] is performed, each search results with correspondence are unified and the depth map was generated, it is effective in lack of information being prevented and a high-precision depth map being obtained.

[0158]The object transfer equipment concerning Claim 10 of this invention, Have three or more image input means, and the with means corresponding to a block chooses two or more kinds of picture pairs which chose the inner 2 \*\* picture of each picture acquired from the three or more above-mentioned image input means, Since the image processing device which performs with [ corresponding to a block ], unifies each search results with correspondence, and generates a depth map should be combined with the object transfer means, an objective position is detected with sufficient accuracy and a transfer can be performed correctly.

[0159]The image processing device concerning Claim 11 of this invention, The depth map creating means which generates the depth map of two or more loaded objects, and a highest rung surface area extraction means to extract the highest rung surface area of the object located in the highest rung from a depth map, The object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from the output of a highest rung surface area extraction means, The object size database which stores the size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of the object candidate combination hypothesis which the above-mentioned hypothesis generating means generated using the information included in a shade original image, It should have the information integration means which recognizes based on a hypothesis with the highest evaluation value among two or more hypotheses, and unifies this recognized picture and the distance information of each object obtained by the depth map creating means. Since the position of two or more candidate objects is detected from the arrangement hypothesis of an object body to the highest rung surface area extracted from the depth map, 1 time may be sufficient as the signal transduction to a transfer means, and hour corresponding is short, and when beforehand based on a texture \*\*\*\* loading method, loading by mistake can also be detected and it can also be determined further whether a highest rung object grasps either.

[0160]The object transfer equipment concerning Claim 12 of this invention, The depth map creating means which generates the depth map of two or more loaded objects, and a highest rung surface area extraction means to extract the highest rung surface area of the object located in the highest rung from a depth map, The object candidate combination hypothesis generating means which carries out two or more listing generation of an object candidate's combination as a hypothesis from the output of a highest rung surface area extraction means, The object size database which stores the size data of recognition objects, a two-dimensional reference pattern, and a hypothetical validation means to verify the validity of the object candidate combination hypothesis which the above-mentioned hypothesis generating means generated using the information included in a shade original image, Recognize based on a hypothesis with the highest evaluation value among two or more hypotheses, and it has an information integration means which unifies this recognized picture and the distance information of each object obtained by the depth map creating means, Since the position of two or more candidate objects is detected from the arrangement hypothesis of an object body to the highest rung surface area which combined with the image processing device and object transfer means which output the three dimensional position information on each object body, and was extracted from the depth map, 1 time may be sufficient as the signal transduction to a transfer means, its hour corresponding is short, and it is also detectable to load by mistake, when beforehand based on a texture \*\*\*\* loading method, and further, It can also be determined whether a highest rung



object grasps either.

[0161]The image processing device concerning Claim 13 of this invention, The object arrangement reference database which stores the data used as the standard of object arrangement, The picture which a hypothetical validation means to verify the validity of the object candidate combination hypothesis which the hypothesis generating means generated has recognized, Since it shall have an object arrangement judging means which judges whether it is right arrangement as compared with criterion data, and a warning generation means which warns a worker of the result of a judgment when the recognized picture does not consistent with criterion data, When [ from which a recognition object stacks and a pattern serves as a standard ] stacking and not having become a pattern, it can warn a device user and the administrator of a host system of not being stacked correctly, and a dangerous state can be avoided.

[0162]The object transfer equipment concerning Claim 14 of this invention, The object arrangement reference database which stores the data used as the standard of object arrangement, The picture which a hypothetical validation means to verify the validity of the object candidate combination hypothesis which the hypothesis generating means generated has recognized, The object arrangement judging means which judges whether it is right arrangement as compared with the above-mentioned criterion data, Since the image processing device and transfer means which have a warning generation means which warns a worker of the result of a judgment should be combined when the recognized picture did not consistent with criterion data, When [ from which a recognition object stacks and a pattern serves as a standard ] stacking and not having become a pattern, it can warn a device user and the administrator of a host system of not being stacked correctly, and a dangerous state can be avoided.

[0163]The image processing device concerning Claim 15 of this invention, The position of the target object is detected from the object size database which stores the data used as the standard of object arrangement using the artificial reference pattern generated automatically, The actual image reference pattern logging means which starts the imaging range which is equivalent to said object body in a detection position, It has an actual image reference pattern storing means which stores the cut-down actual image reference pattern, It constitutes so that said actual image reference pattern may be used in subsequent recognition operation and it may operate, and the thing of small capacity may be sufficient as an image memory, image input time becomes short, and the object of the 2nd henceforth can detect it correctly earlier.

[0164]The object transfer equipment concerning Claim 16 of this invention, The position of the target object is detected from the object size database which stores the data used as the standard of object arrangement using the artificial reference pattern generated automatically, The actual image reference pattern logging means which starts the imaging range which is equivalent to an object body in a detection position, It has an actual image reference pattern storing means which stores the cut-down actual image reference pattern, Since the image processing device and transfer means which were constituted so that an actual image reference pattern might be used in subsequent recognition operation and it might operate should be combined, in image input time, an object serves as short an object transfer equipment which can carry out a detection prickle transfer correctly.

[0165]The image processing device concerning Claim 17 of this invention, The depth map creating means which generates the depth map of two or more tiered object bodies, Based on the information on the depth map stored in the first depth map storing means that stores the generated depth map, and this first depth map creating means, at the time of the initial bundle which calculates a position at the time of an objective bundle A position calculation means, The second depth map storing means that stores the depth map of the remaining cargo after evacuating an object to a position temporarily, A depth map comparison means to compare the depth map stored in the first depth map storing means with the depth map stored in the second depth map storing means, Since it shall have a position-correction-quantity detection means to determine the gripping position of the object evacuated based on the comparison result of this depth map comparison means, and to detect the difference of this gripping position and an initial gripping position as position correction value, Objective position recognition is possible only at a



depth map, it can apply also the object which neither a shade nor edge image can use easily, and processing time also becomes short.

[0166]The object transfer equipment concerning Claim 18 of this invention, The depth map creating means which measures the distance distribution from the image input means of two or more tiered object bodies, The first depth map storing means that stores the generated depth map, and the object transportation device which grasps an object and moves, At the time of the initial bundle which calculates a position based on the information on the above-mentioned depth map at the time of an objective bundle, a position calculation means, The second depth map storing means that stores the depth map of the remaining cargo after grasping an object in a gripping position and making it contrast with a position temporarily by an object transfer means, A depth map comparison means to compare the depth map stored in the first depth map storing means with the depth map stored in the second depth map storing means, The object transfer means which grasps and transfers the image processing device and object which have a position-correction-quantity detection means to determine the gripping position of the object evacuated using the comparison result of a depth map, and to detect a difference with an initial gripping position as position correction value is combined, and objective position recognition can be performed only in a depth map, It becomes an object transfer equipment which is correctly recognized also to the object which neither a shade nor an edge image can use easily, and can be safely transferred to an appropriate position.

---

[Translation done.]

#### \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

#### DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]this embodiment of the invention -- it is a block diagram showing the composition of 1.

[Drawing 2]It is a flow chart which shows the flow of operation of embodiment 1.

[Drawing 3]It is the picture generated in each stage in embodiment 1.

[Drawing 4]this embodiment of the invention -- it is a block diagram showing the composition of 2.

[Drawing 5]It is a flow chart which shows the flow of operation of embodiment 2.

[Drawing 6]It is an exposure explanatory view of a random dot pattern.

[Drawing 7]It is an explanatory view of the block matching of a stereo image.

[Drawing 8]this embodiment of the invention -- it is a block diagram showing the composition of 3.

[Drawing 9]It is a flow chart which shows the flow of operation of embodiment 3.

[Drawing 10]It is an explanatory view of the situation of extracting a level difference region part from a low resolution depth map, accepting the level difference region part of a high resolution depth map, and performing the stereo correspondences of high resolution.

[Drawing 11]this embodiment of the invention -- it is a block diagram showing the composition of 4.

[Drawing 12] It is a flow chart which shows the flow of operation of embodiment 4.

[Drawing 13] this embodiment of the invention -- it is a block diagram showing the composition of 5.

[Drawing 14] It is a flow chart which shows the flow of operation of embodiment 5.

[Drawing 15] It is a flow chart of the with operation corresponding to a stereo block.

[Drawing 16] It is a figure showing a situation with the stereo correspondences of a stereo image typically.

[Drawing 17] this embodiment of the invention -- it is a block diagram showing the composition of 6.

[Drawing 18] It is a flow chart which shows the flow of operation of embodiment 6.

[Drawing 19] It is a figure showing each \*\*\*\*\* of an operation situation which extracts an object candidate using an object candidate hypothesis.

[Drawing 20] this embodiment of the invention -- it is a block diagram showing the composition of 7.

[Drawing 21] It is a flow chart which shows the flow of operation of embodiment 7.

[Drawing 22] this embodiment of the invention -- it is a block diagram showing the composition of 8.

[Drawing 23] It is a flow chart which shows the flow of operation of embodiment 8.

[Drawing 24] this embodiment of the invention -- it is a block diagram showing the composition of 9.

[Drawing 25] It is a flow chart which shows the flow of operation of embodiment 9.

[Drawing 26] It is a figure showing each stage of the operation which amends an initial gripping position.

[Drawing 27] It is an explanatory view showing the procedure of the conventional image processing device.

[Drawing 28] It is a flow chart of operation of the processing shown in drawing 27.

[Drawing 29] It is a figure showing the principle of spatial-code-izing using pattern light.

[Drawing 30] It is an explanatory view of the depth map instrumentation system of other conventional examples.

[Drawing 31] It is an explanatory view of the stereo-correspondences searching device of another conventional Image Processing Division.

[Drawing 32] It is a flow chart of another conventional Image Processing Division drawing 30.

[Description of Notations]

11 An image input means and 12 A depth map creating means and 13 Highest rung surface area extraction means, 14 A highest rung object candidate extraction means and 15 A shade original image storing means and 16 Object size database, 17 A two-dimensional reference pattern creating means and 18 An object position detecting means and 19 Information integration means, 20 A random texture pattern light projection means and 21 Image input means, 22 The with means corresponding to a stereo image block, and 30 A pattern floodlight and 31 The with means corresponding to a low resolution block, 32 A level difference region extraction means, the with means corresponding to 33 high-resolution block, and 34 Depth map synthesizing means, 41a The first image storing means and 41b The second image storing means and 42 High-speed similarity calculation means, 43 A high reliability similarity calculation means, 44 corresponding-block reliability judgment means, and 45 Depth map storing means, 51a The first image input means and 51b The second image input means and 51c The third image input means, 52 The with means corresponding to a stereo image block, and 64 Object candidate combination hypothesis generating means, 65 A hypothetical validation means and 69 [ A warning generation means and 81 / An artificial reference pattern creating means and 82 / Actual image reference pattern logging means, ] An object position detecting means and 71 An object arrangement reference database and 72 An object arrangement judging means and 73 83 An actual image reference pattern storing means and 88 [ The second depth map storing means and 93 / A depth map comparison means and 94 / An initial gripping position calculating means and 95 / A position-correction-quantity calculating means and 96 / Object transportation device. ] An object position detecting means and 91 A depth map creating means and 92a The first depth

[Translation done.]

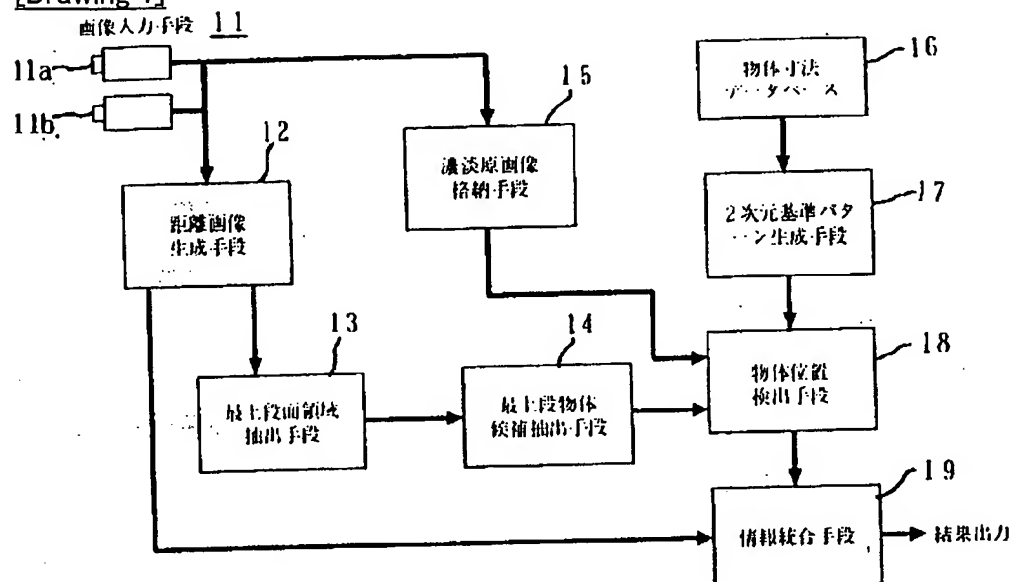
\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

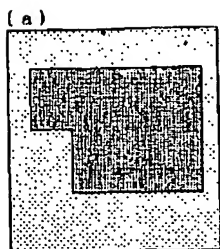
DRAWINGS

[Drawing 1]

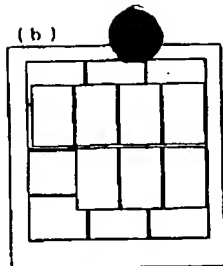


- |                    |                  |
|--------------------|------------------|
| 11 : 画像入力手段        | 12 : 距離画像生成手段、   |
| 13 : 最上段面領域抽出手段    | 14 : 最上段物体候補抽出手段 |
| 15 : 濃淡原画像格納手段     | 16 : 物体寸法データベース  |
| 17 : 2次元基準パターン生成手段 | 18 : 物体位置検出手段    |
| 19 : 情報統合手段        |                  |

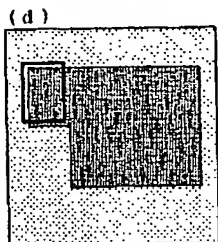
[Drawing 3]



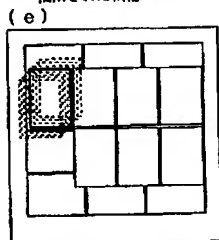
抽出された物体領域



原画像



抽出された候補の一つ



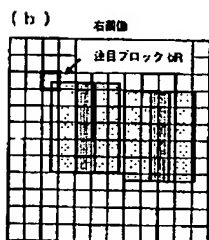
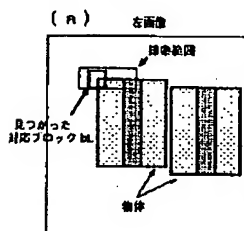
位置決めされた物体

(c)

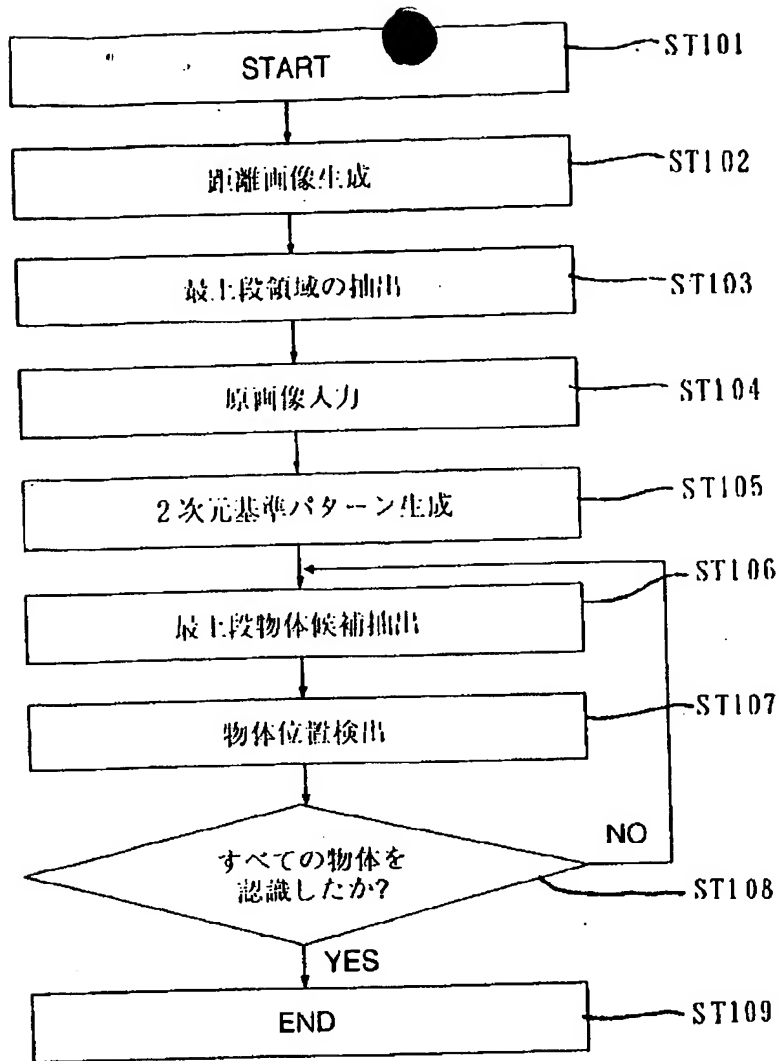


物体の寸法データを用いて  
自動的に生成されたテンプレート

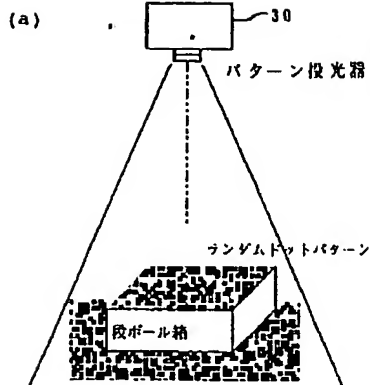
## [Drawing 7]



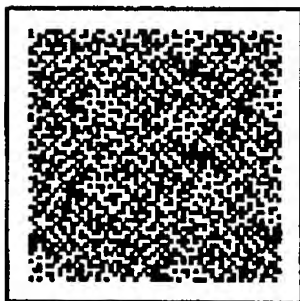
## [Drawing 2]



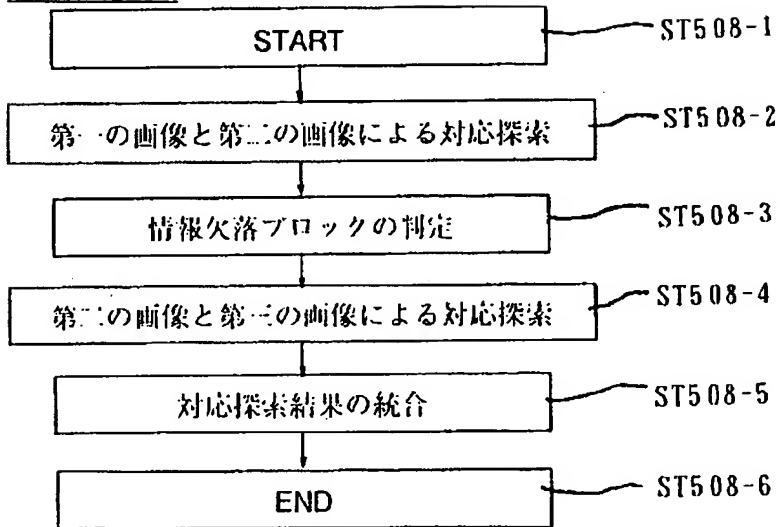
[Drawing 6]



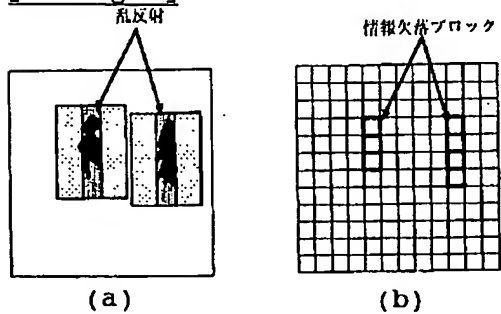
(b)



[Drawing 15]



[Drawing 16]





[Drawing 4]

ランダムテクスチャ

パターン投光手段



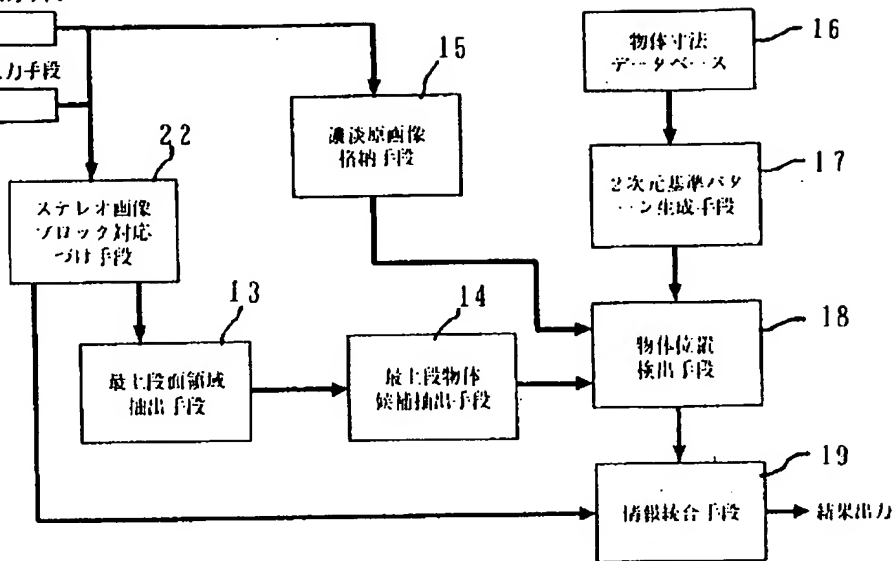
20

第一の画像入力手段

21a

第二の画像入力手段

21b



20 : ランダムテクスチャパターン投光手段

21 : 画像入力手段

22 : ステレオ画像ブロック対応付手段

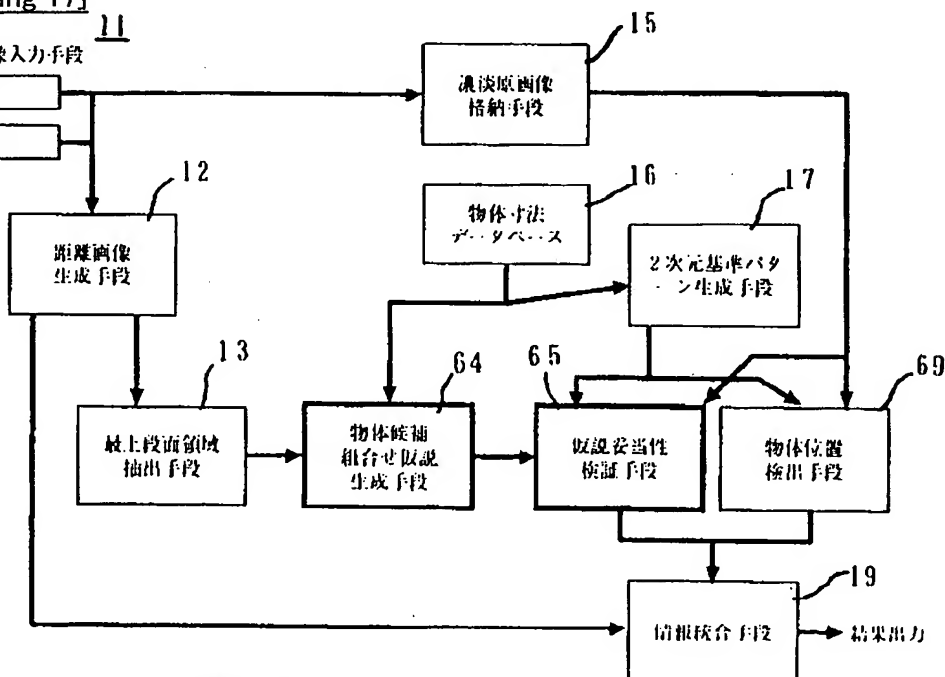
[Drawing 17]

11

画像入力手段

11a

11b

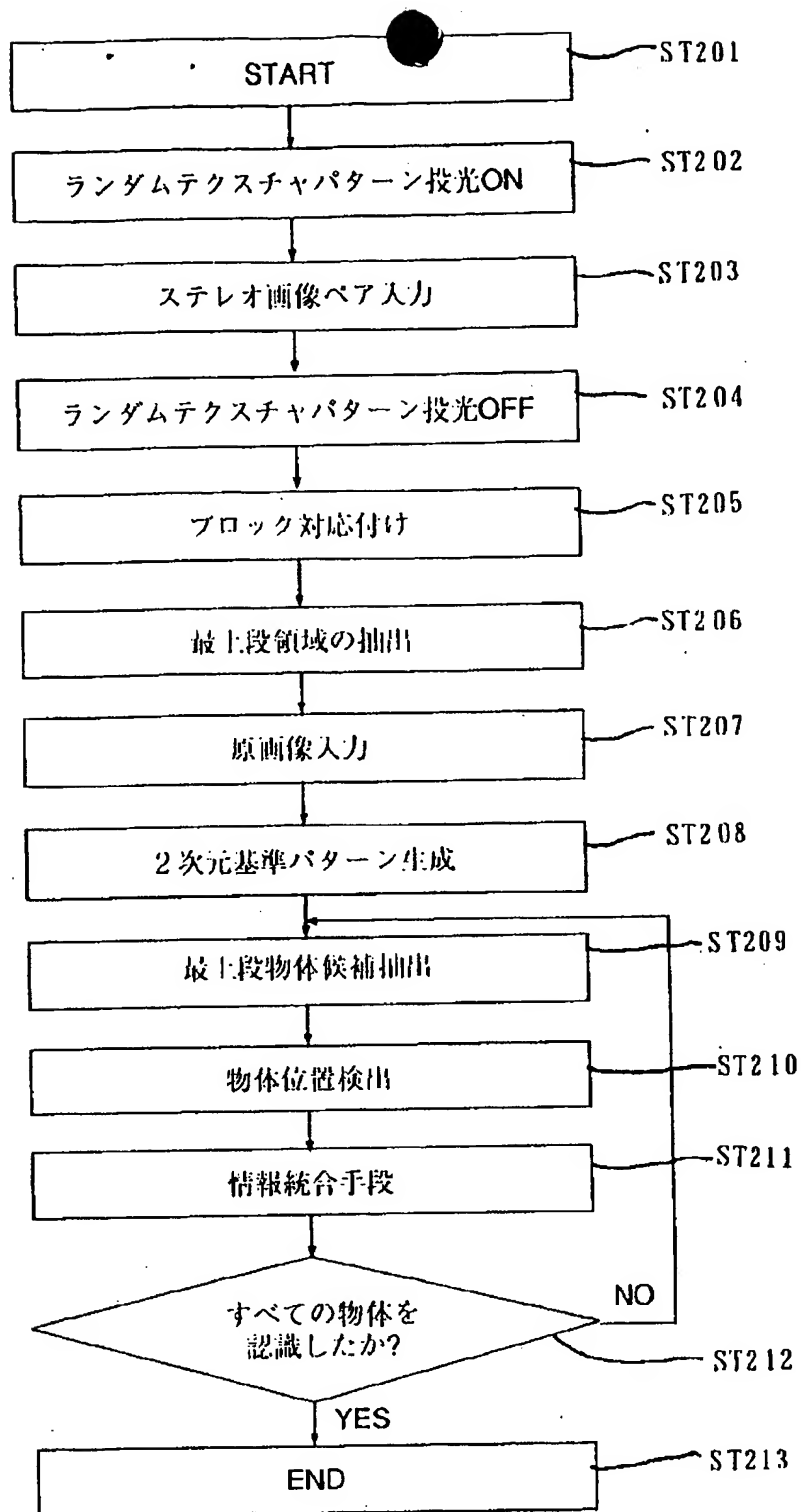


64 : 物体候補組合せ仮説生成手段

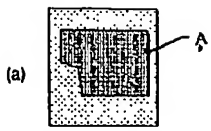
65 : 仮説妥当性検証手段

69 : 物体位置検出手段

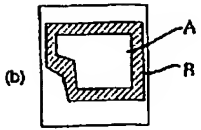
[Drawing 5]



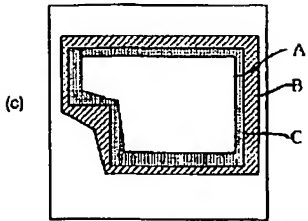
[Drawing 10]



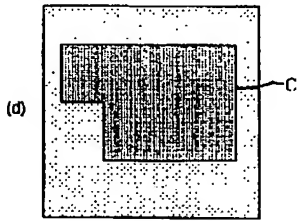
低解像度距離画像



段差領域部

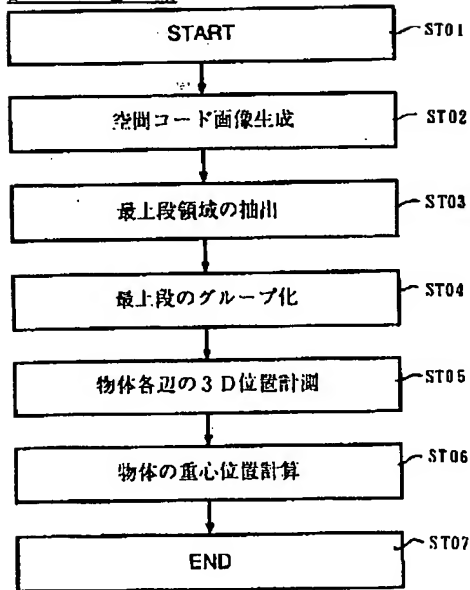


段差領域に対応する高解像度距離画像



合成された距離画像

[Drawing 28]



[Drawing 8]

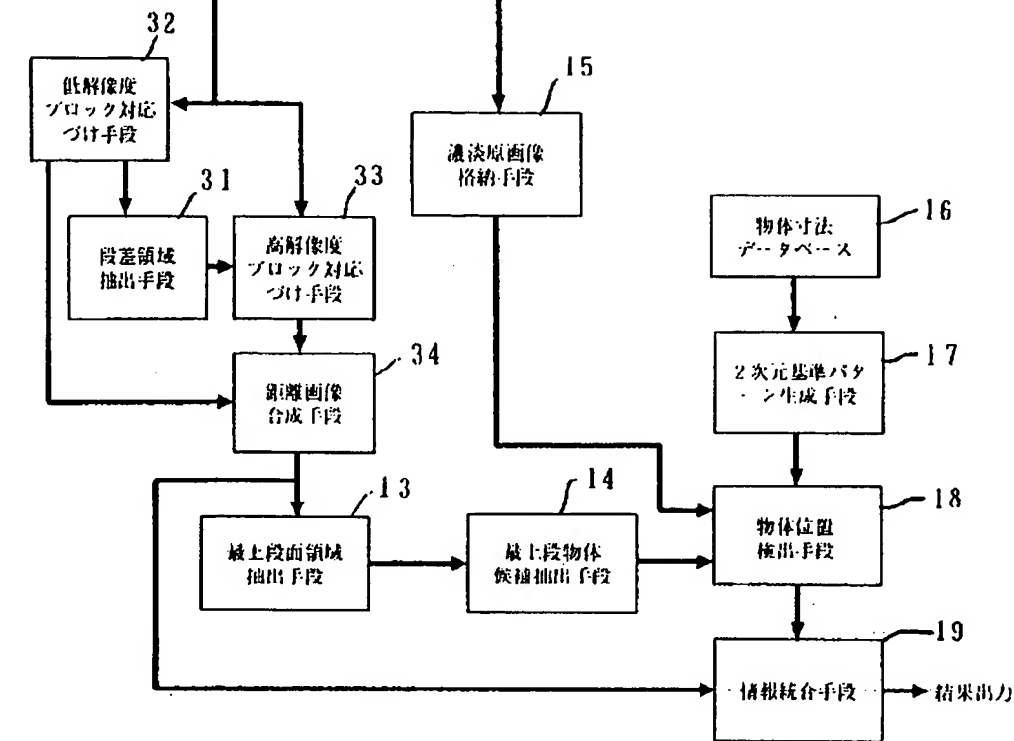
ランダムテクスチャ  
パターン投光手段、  
20

第一の画像入力手段

21a

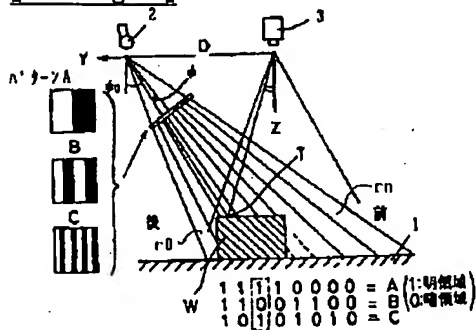
第二の画像入力手段

21b



- 31 : 低解像度ブロック対応付手段  
 32 : 段差領域抽出手段  
 33 : 高解像度ブロック対応付手段、  
 34 : 距離画像合成手段

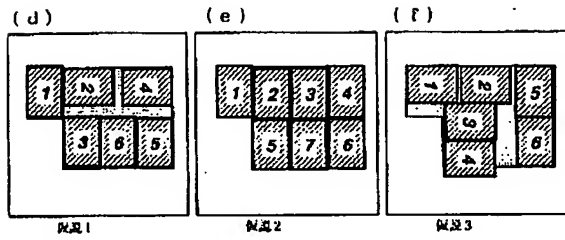
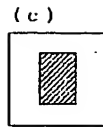
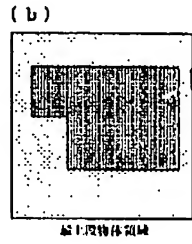
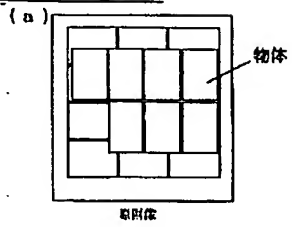
[Drawing 29]



[Drawing 30]

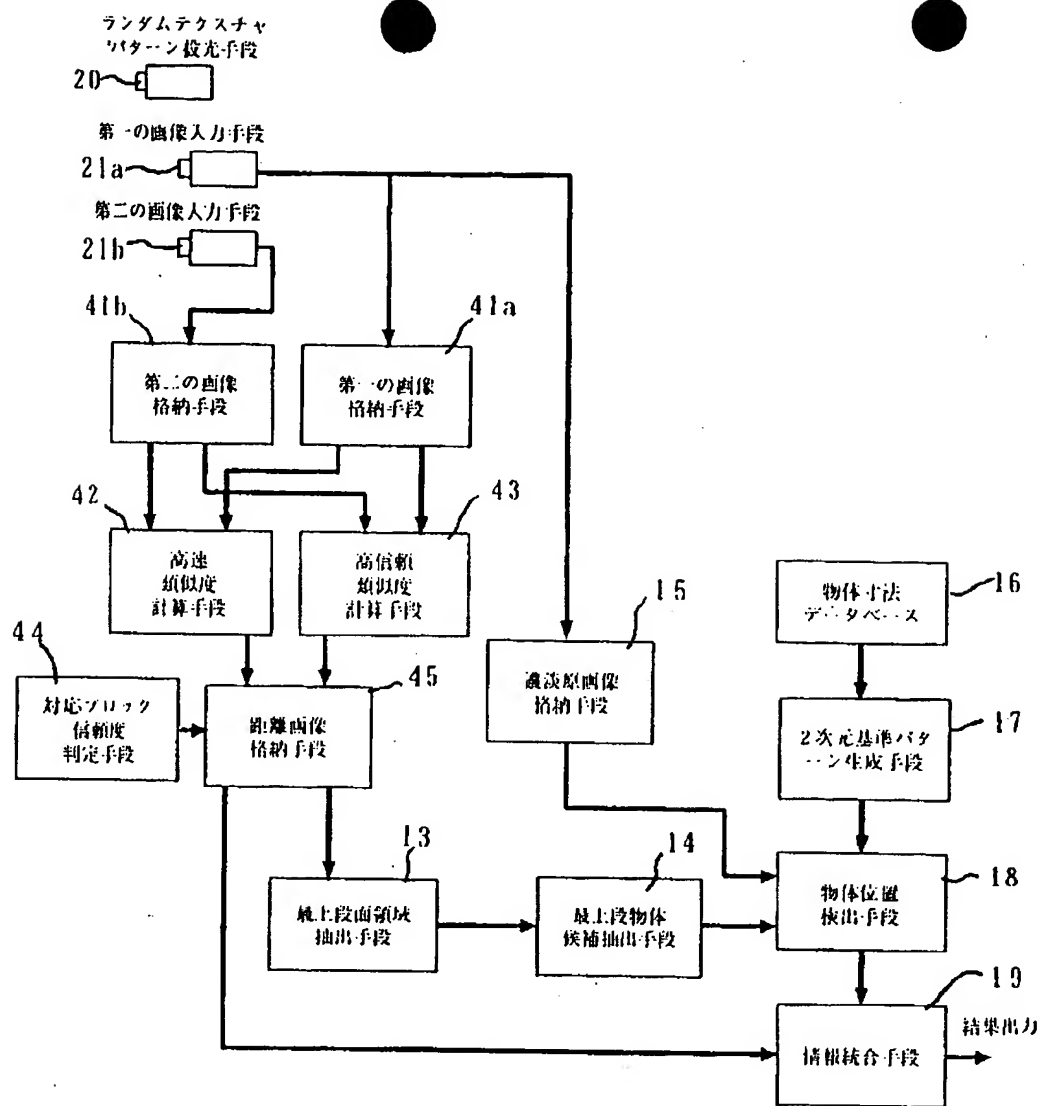


# [Drawing 19]



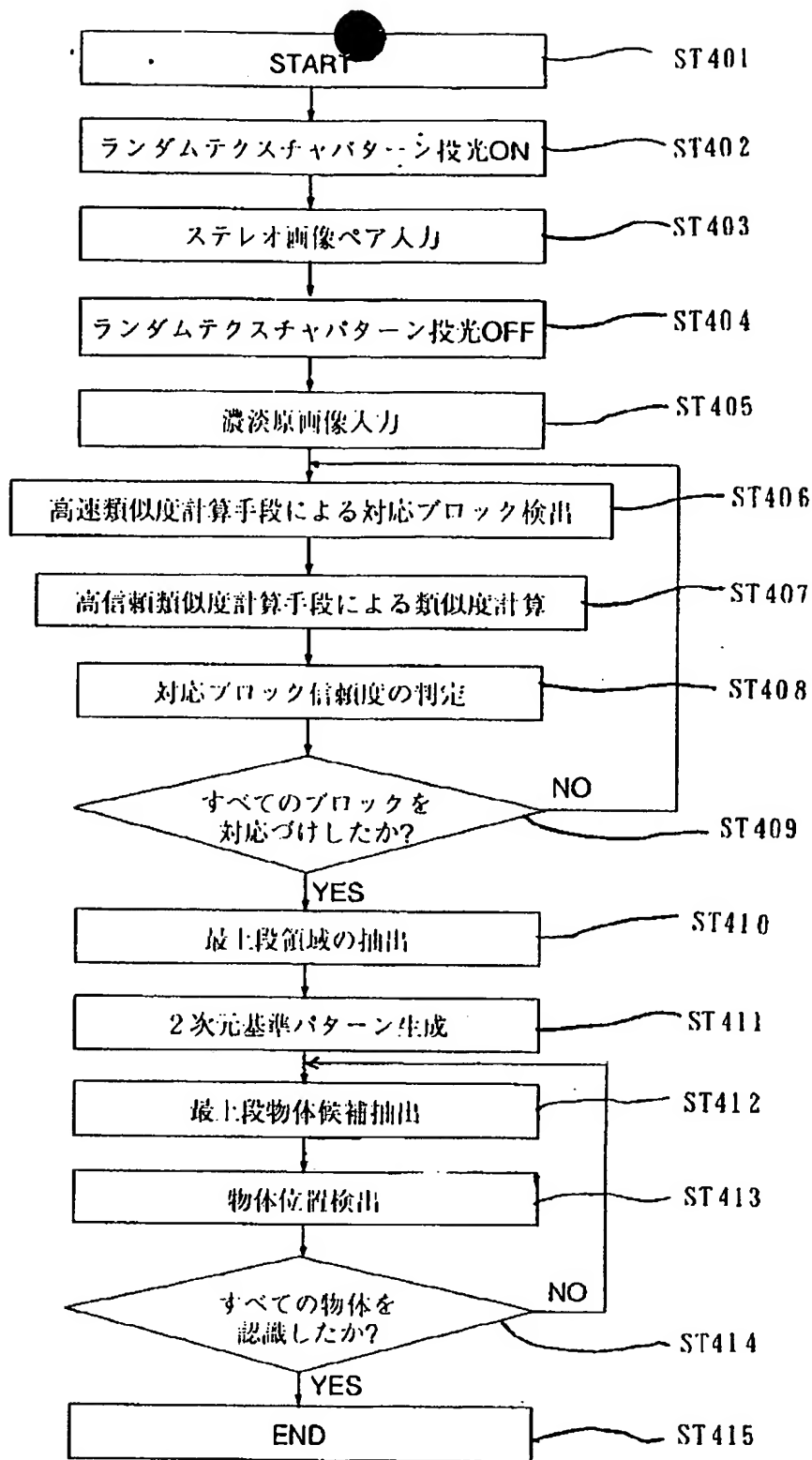
# [Drawing 11]



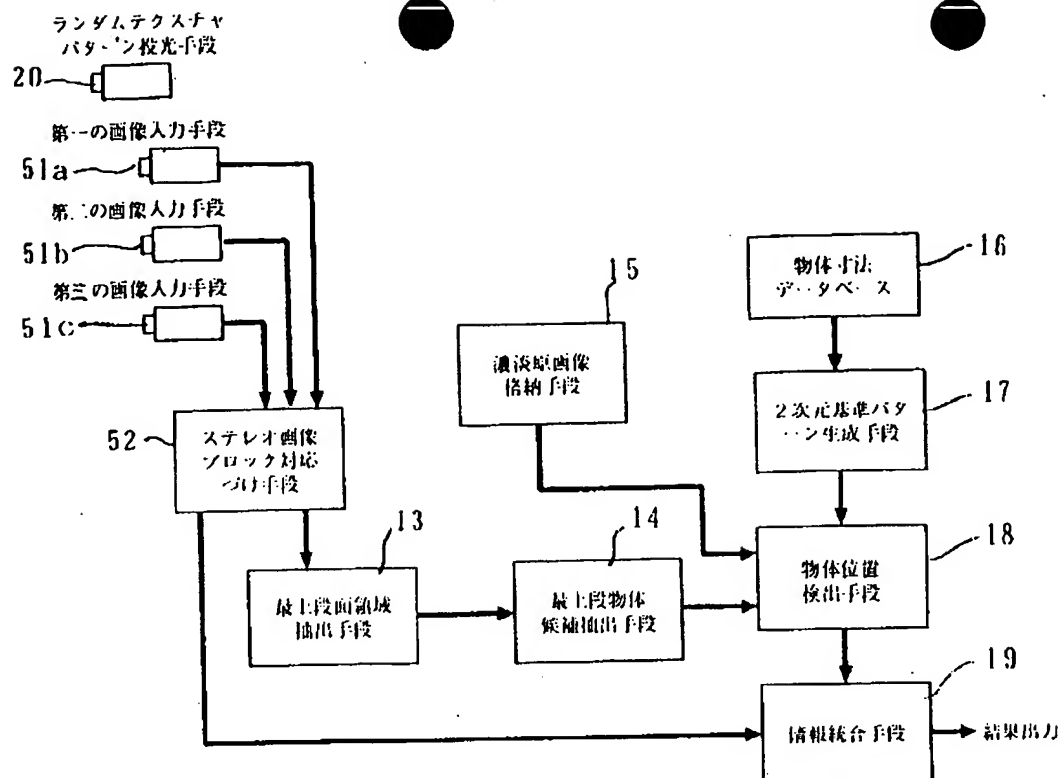


- 41a : 第一の画像格納手段  
 41b : 第二の画像格納手段  
 42 : 高速類似度計算手段  
 43 : 高信頼類似度計算手段  
 44 : 対応ブロック信頼度判定手段  
 45 : 距離画像格納手段

[Drawing 12]

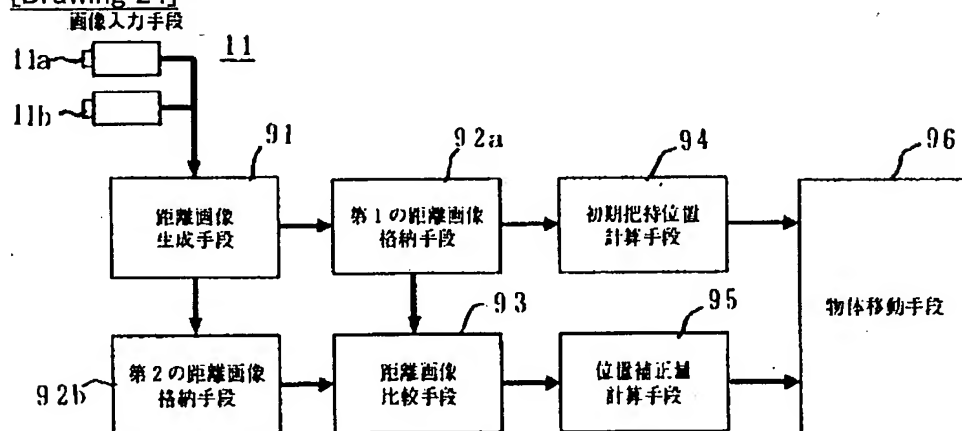


[Drawing 13]



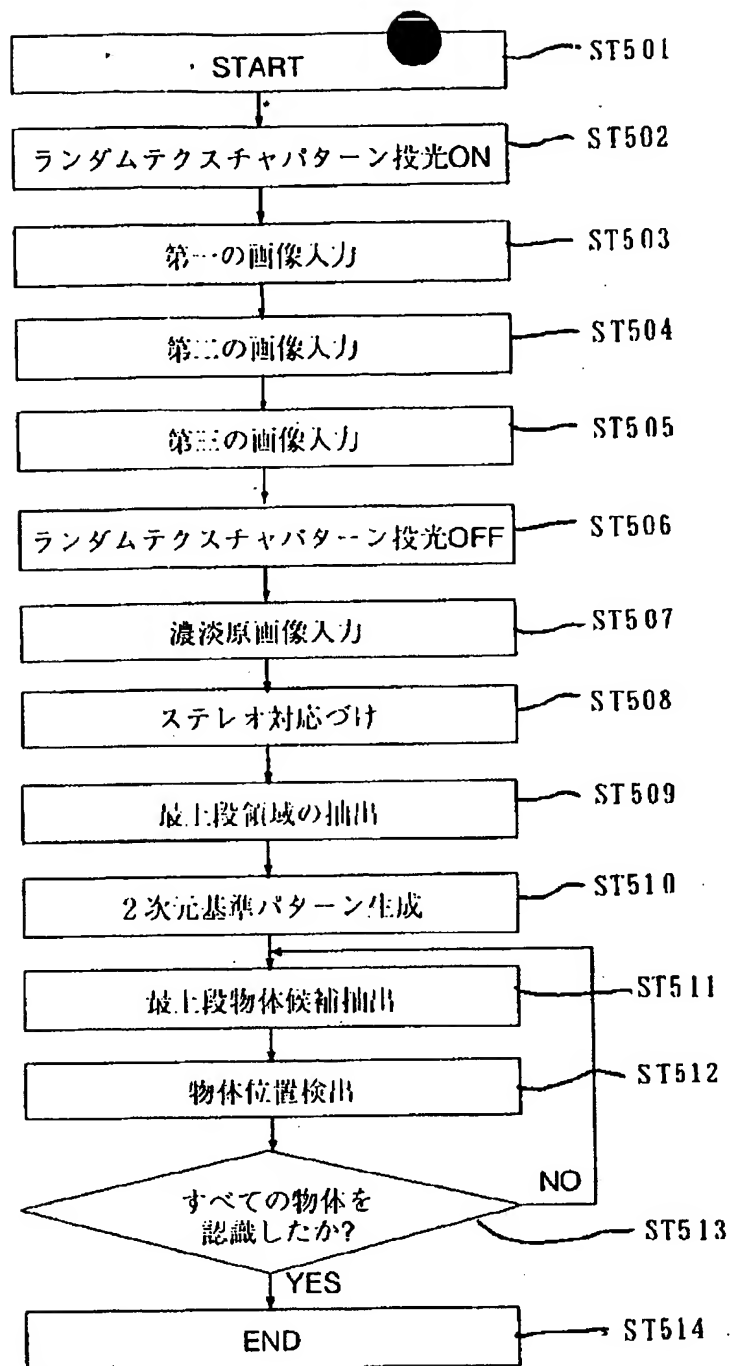
51a : 第一の画像入力手段  
51b : 第二の画像入力手段  
51c : 第三の画像入力手段  
52 : ステレオ画像ブロック対応付手段

[Drawing 24]

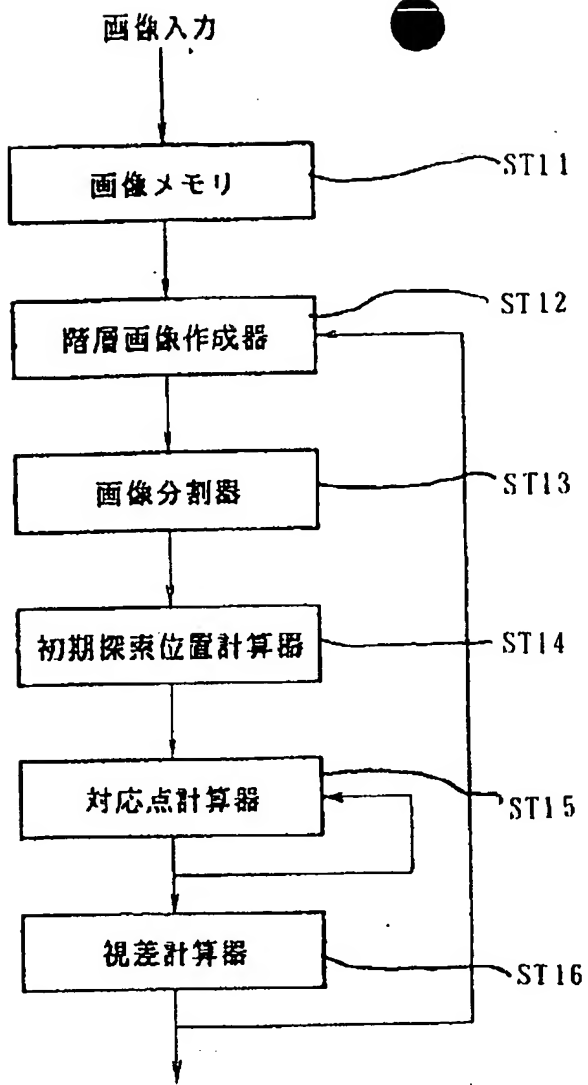


91 : 距離画像生成手段  
92a : 第一の距離画像格納手段  
92b : 第二の距離画像格納手段  
93 : 距離画像比較手段  
94 : 初期把持位置計算手段  
95 : 位置補正量計算手段  
96 : 物体移動手段

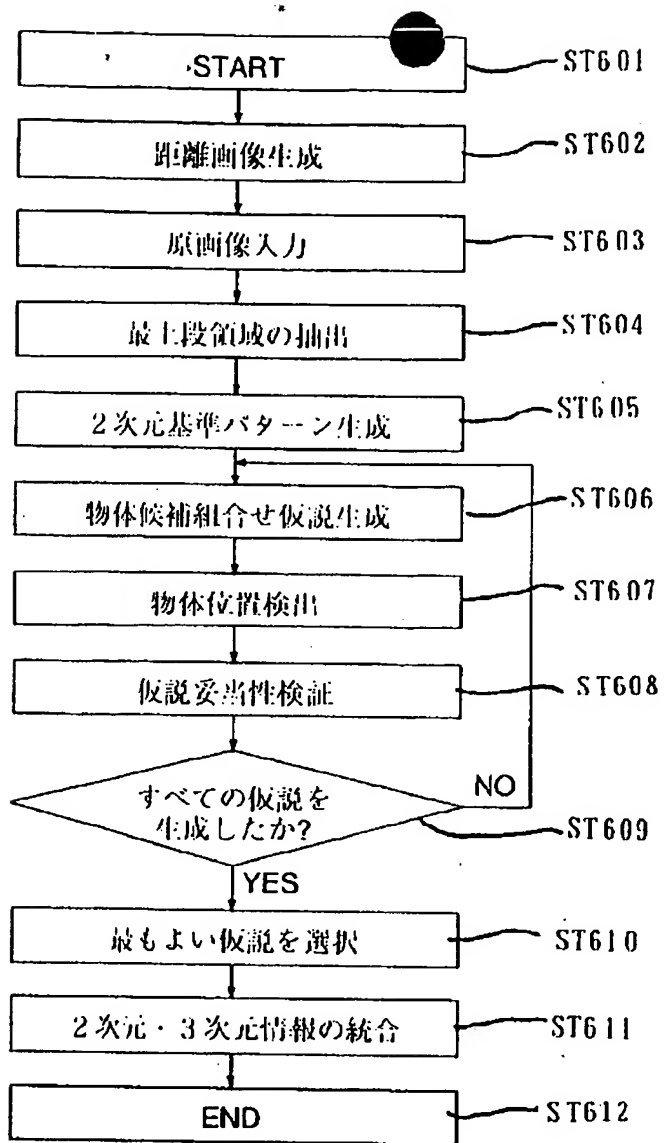
[Drawing 14]



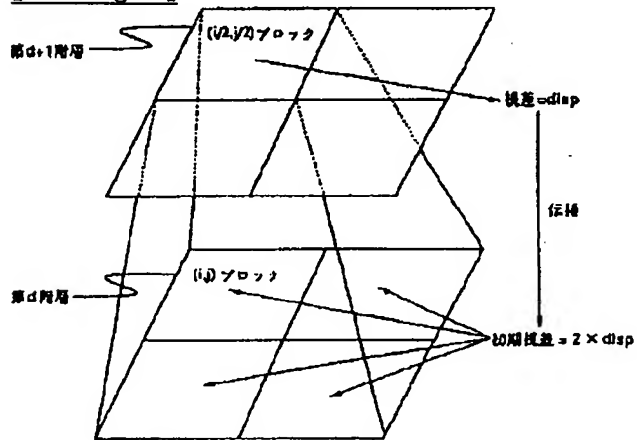
[Drawing 32]



[Drawing 18]

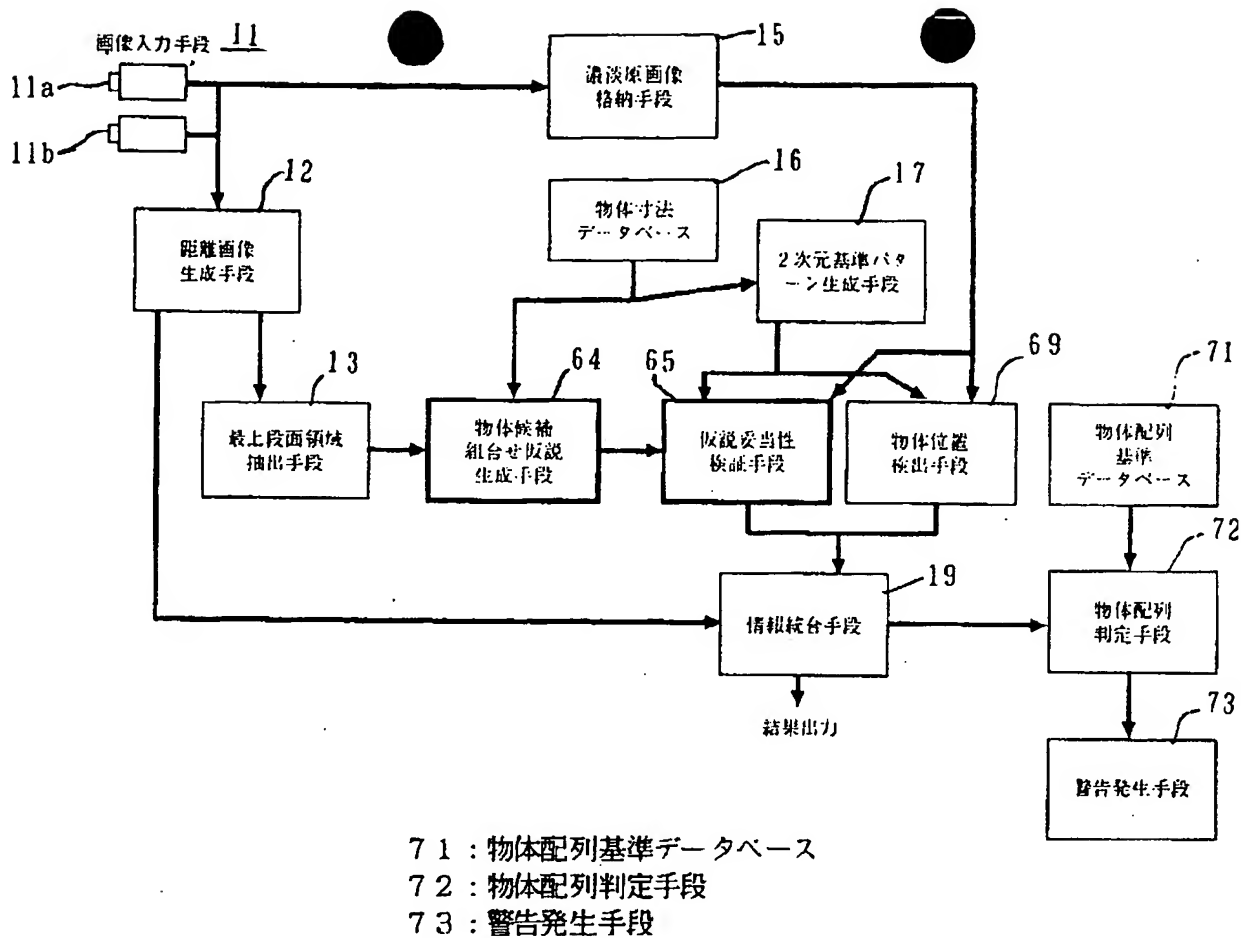


[Drawing 31]

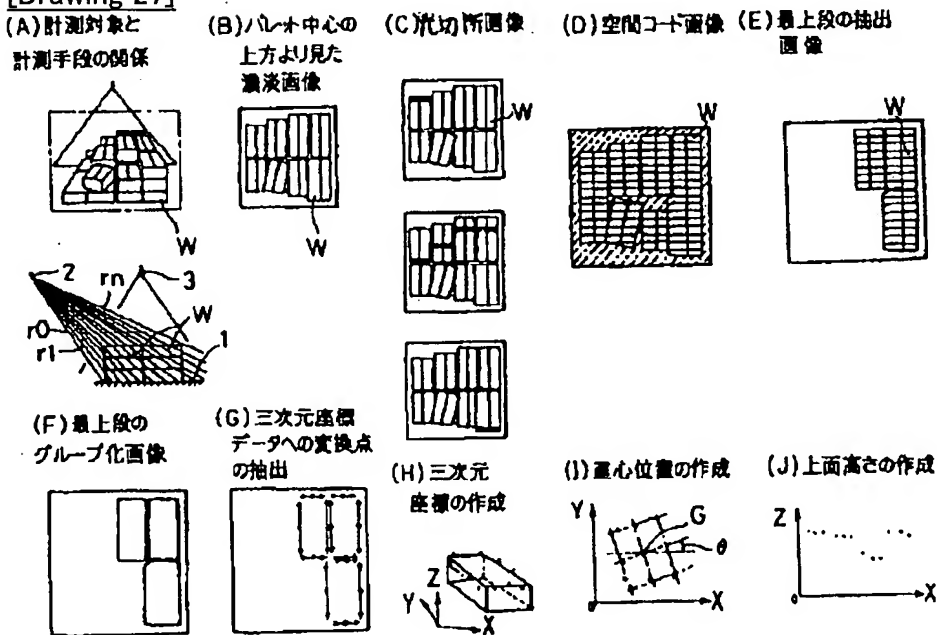


[Drawing 20]

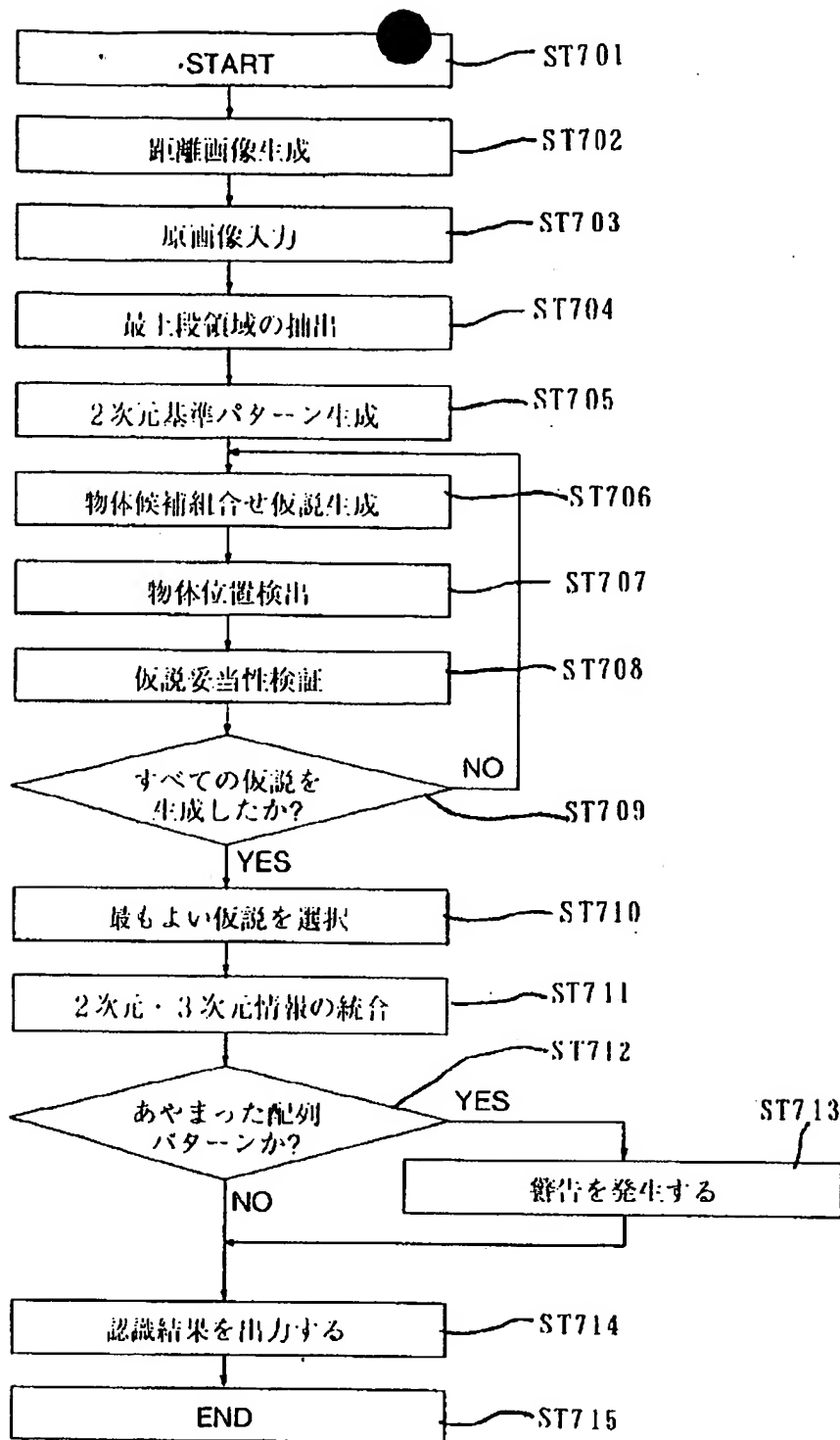




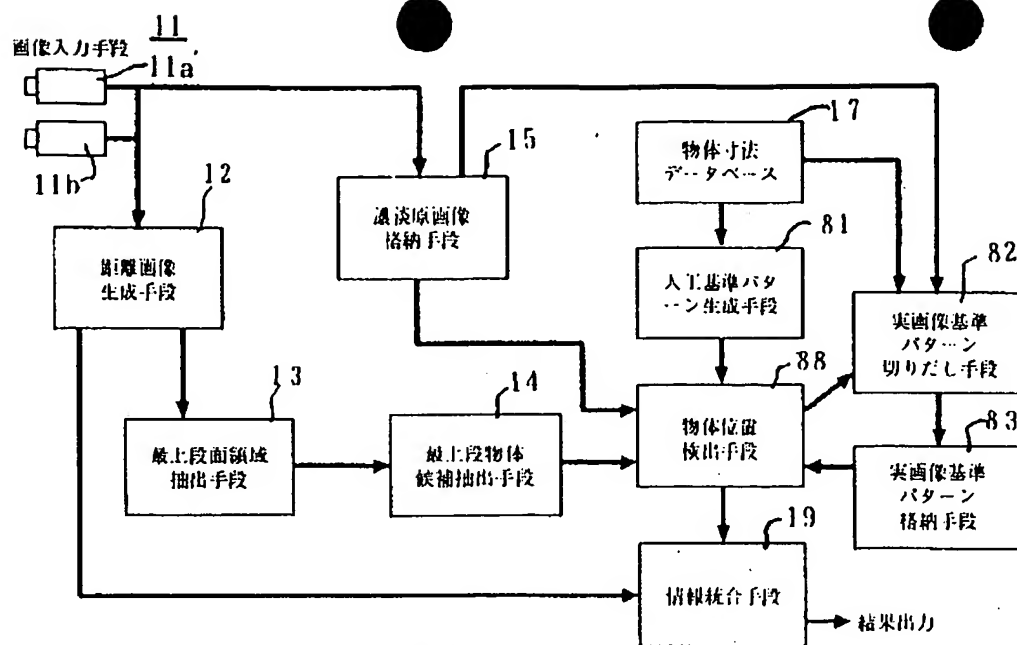
[Drawing 27]



[Drawing 21]

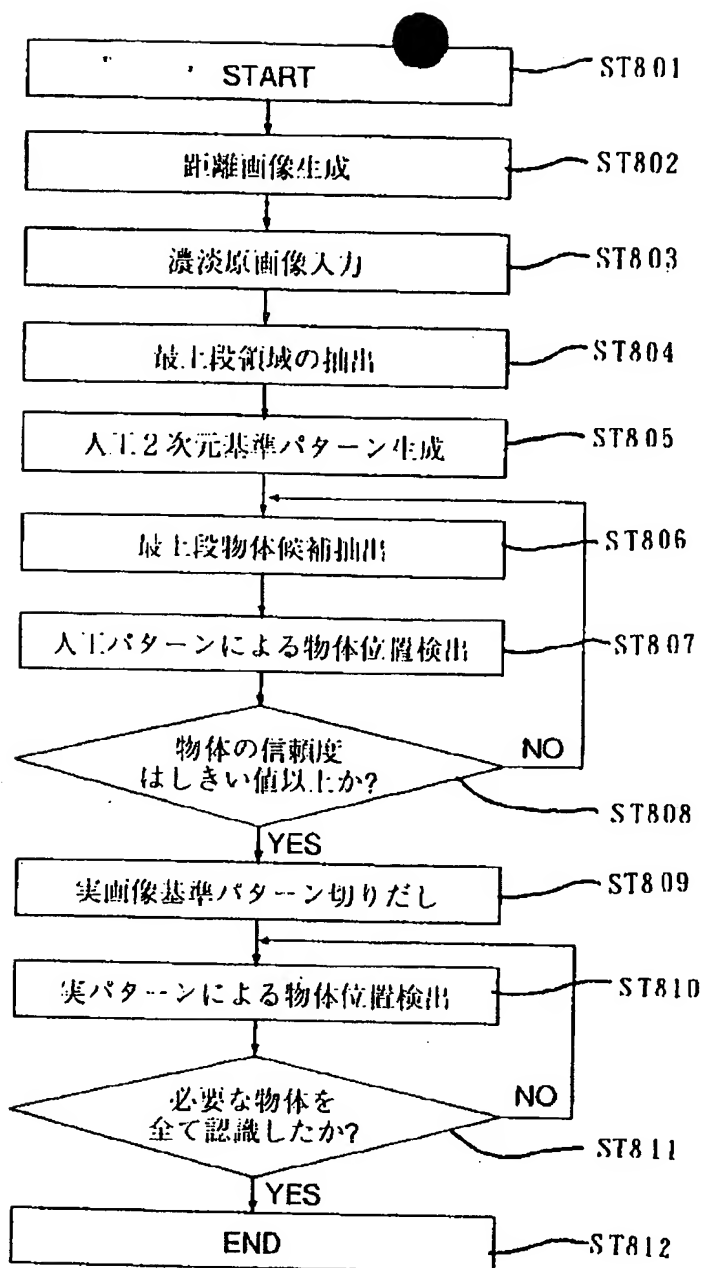


[Drawing 22]

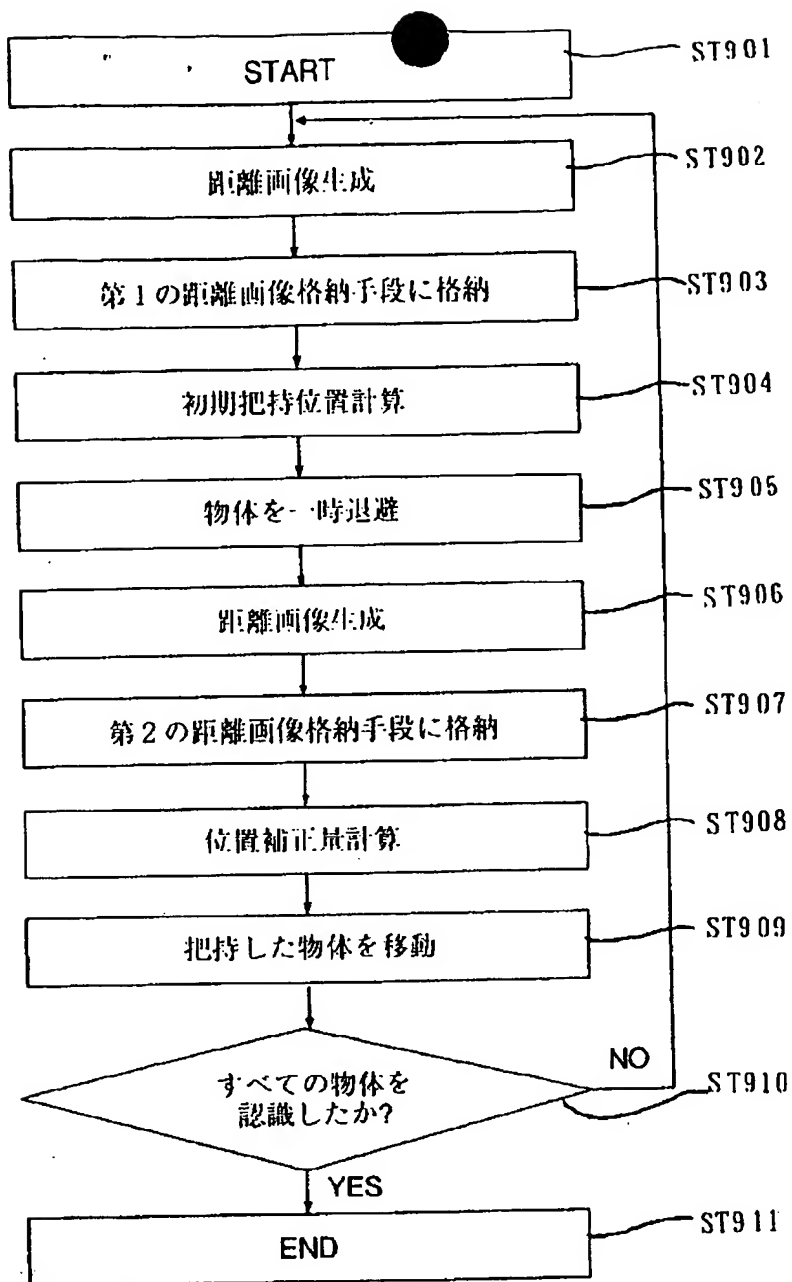


- 81 : 人工基準パターン生成手段  
 82 : 実画像基準パターン切り出し手段  
 83 : 実画像基準パターン格納手段  
 88 : 物体位置検出手段

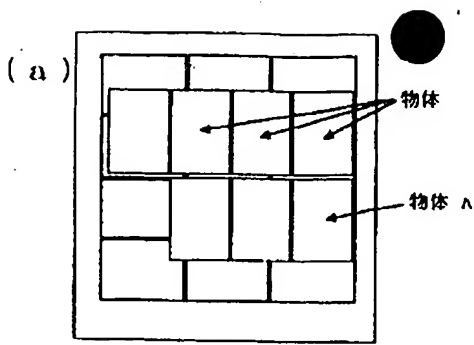
[Drawing 23]



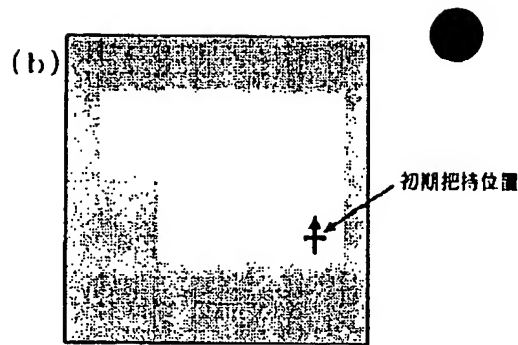
[Drawing 25]



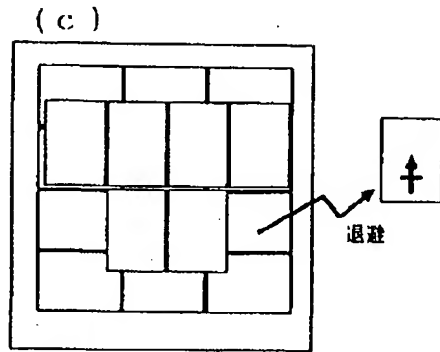
[Drawing 26]



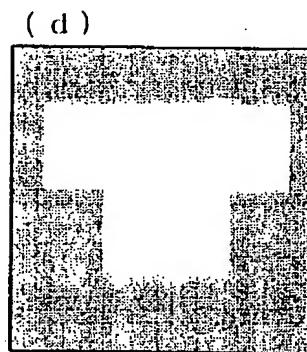
物体退避前の原画像



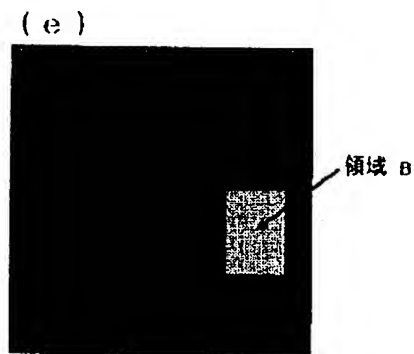
第1の距離画像



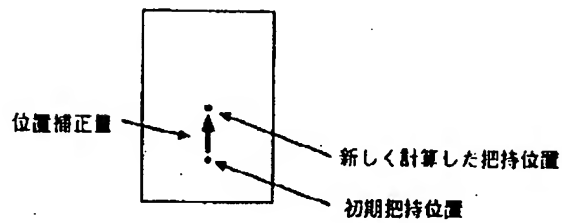
物体退避後の原画像



第2の距離画像



距離画像の差分結果



(f) 退避した物体の拡大図

[Translation done.]



特開平10-31742

(43) 公開日 平成10年(1998) 2月3日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 T 7/00			G 0 6 F 15/62	4 1 5
G 0 1 B 11/00			G 0 1 B 11/00	H
G 0 6 T 1/00			G 0 6 F 15/64	M
7/60			15/70	3 5 0 B

審査請求 未請求 請求項の数18 O L (全 40 頁)

(21) 出願番号 特願平8-184917

(22) 出願日 平成8年(1996) 7月15日

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72) 発明者 橋本 学

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(72) 発明者 鷺見 和彦

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(74) 代理人 弁理士 宮田 金雄 (外3名)

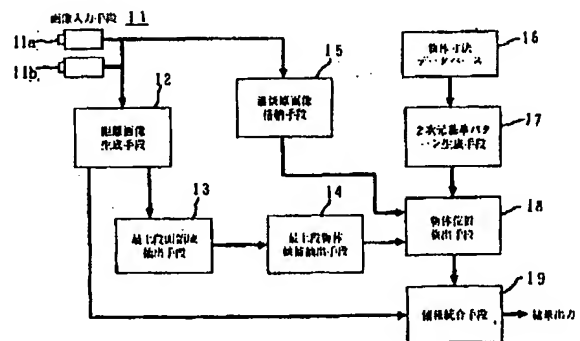
## (54) 【発明の名称】 画像処理装置および物体移載装置

## (57) 【要約】

【課題】 積荷を自動的に移載するロボットや産業機械の積載された荷物の位置を認識する画像処理装置及びその装置を組み込んだ物体移載装置の物体の位置の認識を素早く正確に行えるものとする。

【解決手段】 最上段領域の抽出は、粗い画像の距離画像で行い、物体の位置検出は物体のデータベースを用いて2次元基準パターンを生成し、最上段領域にパターンマッチングして物体位置を認識するようにした。

【効果】 最上段領域の認識は粗い画像の距離画像としたことにより、画像メモリが少なくなつて、画像入力時間が早くなり、物体の位置検出は物体の寸法データより生成した基準パターンを用いて、正確な位置が素早く検出できる画像処理装置及び物体移載装置が得られる。



- |                    |                  |
|--------------------|------------------|
| 11 : 画像入力手段        | 12 : 距離画像生成手段、   |
| 13 : 最上段領域抽出手段     | 14 : 最上段物体位置検出手段 |
| 15 : 薄化原画像格納手段     | 16 : 物体寸法データベース  |
| 17 : 2次元基準パターン生成手段 | 18 : 物体位置検出手段    |
| 19 : 情報統合手段        |                  |

## 【特許請求の範囲】

【請求項1】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、上記個別に抽出された最上段物体候補について、上記2次元基準パターンと上記濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、上記各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置。

【請求項2】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、上記個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、上記物体の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置および物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する物体移載装置。

【請求項3】 距離画像生成手段は、ランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、上記第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段とで構成されていることを特徴とする請求項1記載の画像処理装置。

【請求項4】 距離画像生成手段が、ランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、上記第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段で構成された画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2記載の物体移載装置。

【請求項5】 距離画像生成手段は、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、上記低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されていることを特徴とする請求項1記載の画像処理装置。

【請求項6】 距離画像生成手段は、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、上記低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段で構成された画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2記載の物体移載装置。

【請求項7】 ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックのペアに関して高信頼に類似度を計算する高信頼類似度計算手段と、上記高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成された請求項2記載の画像処理装置。

【請求項8】 ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックのペアに関して高信頼に類似度を計算する高信頼類似度計算手段と、上記高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成された画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項4記載の物体移載装置。

【請求項9】 3つ以上の画像入力手段を有し、ブロック対応付手段が上記3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成することを特徴とする請求項1または請求項3記載の画像処理装置。

【請求項10】 3つ以上の画像入力手段を有し、ブロック対応付手段が上記3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成する画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2または請求項4記載の物体移載装置。

【請求項11】 積載された複数の物体の距離画像を生

成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置。

【請求項12】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段を備え、個々の対象物体の三次元位置情報を出力する画像処理装置と物体を把持し移載する移載手段を備え、上記画像処理装置の出力信号に基づいて物体を移載する物体移載装置。

【請求項13】 物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有することを特徴とする請求項11記載の画像処理装置。

【請求項14】 物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と物体を把持し移載する移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項12記載の物体移載装置。

【請求項15】 物体寸法データベースから自動的に生成される人工基準パターンを用い、対象となる物体の位置を検出して、検出位置において上記対象物体に相当す

る画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては上記実画像基準パターンを利用するように動作することを特徴とする請求項1記載の画像処理装置。

【請求項16】 物体寸法データベースから自動的に生成される人工基準パターンを用い、対象となる物体の位置を検出して、検出位置において上記対象物体に相当する画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては上記実画像基準パターンを利用するように動作する画像処理装置と物体を把持し移載する移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2記載の物体移載装置。

【請求項17】 段積みされた複数の対象物体の画像入力手段からの距離分布を計測する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、この第一の距離画像生成手段に格納された距離画像の情報をもとに、物体の把持位置を計算する初期把持位置計算手段と、一時的に物体を所定の位置に退避させた後に、残りの積荷の距離画像を格納する第二の距離画像格納手段と、上記第一の距離画像格納手段に格納された距離画像と、上記第二の距離画像格納手段に格納された距離画像とを比較する距離画像比較手段と、この距離画像比較手段の比較結果を基に上記退避した物体の把持位置を決定し、この把持位置と上記初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有することを特徴とする画像処理装置。

【請求項18】 段積みされた複数の対象物体の画像入力手段からの距離分布を計測する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、物体を把持し移動する物体移動手段と、上記距離画像の情報をもとにして物体の把持位置を計算する初期把持位置計算手段と、上記物体移動手段によって物体を上記把持位置で把持し、一時的に所定の位置に対比させた後に残りの積荷の距離画像を格納する第二の距離画像格納手段と、上記第一の距離画像格納手段に格納された距離画像と、上記第二の距離画像格納手段に格納された距離画像を比較する距離画像比較手段と、上記距離画像の比較結果を使用して上記退避した物体の把持位置を決定し上記初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有する画像処理装置および物体の移載装置を備えた画像処理装置の出力に基づいて、物体移載手段によって物体を移載する物体移載装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、積荷を自動的に移載するロボットや産業機械の積載された荷物の位置を認識する画像処理装置に関するものである。

## 【0002】

【従来の技術】積荷を自動的に移載する手段の画像処理装置の例として特開平6-249631号公報に開示されたものがある。図27はその発明の画像計測の処理手順の概略説明図、図28に動作を示すフローチャート、図29はパターン光を用いた空間コード化の原理を示す図である。

【0003】この画像処理装置は、パレット1上に積載された段ボール箱Wを計測対象とするものであり、図29に示すように、パレット1の斜め上方に投光器2、パレット1の中心部上方に配置されたカメラ3からなり、投光器2は各種のパターン光を時系列的に投光することにより、測定空間を互いに積層された楔状測定領域 $r_0 \sim r_n$ に分割するスリットパターン光を投光するものである。スリットパターンは、例えば図29のA、B、Cの白黒の2値パターンの3つのパターン光を投光するようになっており、白は光の当たった部分、黒は光の当たっていない部分である。このようなパターンは、例えば液晶シャッタのようなドットマトリクス電気シャッタを用いて作ることができる。

【0004】光が当たっている状態を“1”、当たっていない状態を“0”と表現すると、図29においては、パターンAが投光されている時は前から後ろに向かって、半分づつ1、0となる。パターンBでは、前から順に1/4づつ1、0、1、0となる。パターンCでは、1/8づつ1、0、1、0、1、0、1、0となる。

【0005】これら3つのパターンを投光してそれぞれ画像を撮像すると、各画素については「000」から

「111」の3ビットコードにコード化されたデータのどれかに対応させることができる。このようなコードを空間コードと呼ぶ。ある空間コードに対応するのは3次元空間上では楔状領域であり、そこに物体Wが存在していると物体の表面上の画像領域にコードが割り当てられることになる。一方、カメラで撮像した画像から、画像上のある画素の空間コードがわかれば、どの楔状領域に相当するかがわかるので、三角測量の原理によって物体表面の高さ、即ちカメラからの距離がわかることになる。この説明ではパターンの種類をA、B、Cの3種類としたため、画像全体を8つの楔状領域に分割したが、パターンの種類を8種類とすれば、楔状領域の数は256になり、より精密に距離情報の精度を向上させることができる。即ち、空間コード画像を生成することによって、画像上のそれぞれの画素のカメラからの距離が計測できる。

【0006】図27について、フローチャートにしたがって説明する。ステップST01でシステムが起動されると、ステップST02において空間コード画像が生成され、図27(A)の状態にある積荷の上面に同図

(D)のような空間コード画像が得られ、画像を水平に左から右に走査し、もっとも空間コードの大きいデータ

を持つ領域を抽出すれば同図(E)のように最上段物体の上面部分荷対応する画像が得られる。ステップST04では、最上段物体領域をグループ化している。空間コード画像において隣接する画素間で空間コードが所定の範囲内に近接している時、それら領域を同じグループと判断したり、各積荷上縁間に生じる溝状の部分やすき間部分の空間コードの変化をもとにグループ化している。グループ化された各物体は、続くステップST05で各物体の3次元位置が計測される。グループ化された物体の辺縁部分に相当する部分の3次元座標を何点かサンプリングし、ステップST06で、それらをもとに物体の重心位置と姿勢、および高さ情報を得る。ステップST07で処理を終了する。このようにして得られた積荷の位置データを用いて、ロボット等で積荷をひとつづつ移載することができる。

【0007】この画像処理装置では、積荷の認識に必要な情報をすべて空間コード画像から獲得し、例えば、解像度 $256 \times 256$ または $512 \times 512$ 画素といった高い解像度の距離画像を生成する必要がある、時系列的に投光パターンを変化させる機構を有する高精度のパターン投光器が使用され、時系列で発生させた複数の投光パターンそれぞれに対応させて、画像を入力するので、大容量の画像メモリを必要とする。

【0008】他の従来例として、電子情報通信学会論文誌Vol. J71-D, No. 7 p1240~1257に掲載された佐藤、井口の論文「液晶レンジファインダ・・液晶シャッタによる高速距離画像計測システム・・」に記載された内容を紹介する。この技術は、上記の図27~図29に示された技術で使用された空間コード化法による距離画像獲得システムに関するものであり、図30にその原理を示す。液晶シャッタ41を用いて時系列的に交番2進コードによるスリットパターンを物体に投影し、カメラ3でそれぞれのパターンに対応する画像を撮像して空間コード画像を生成するものである。パターンとしては図42のような7枚のパターンを用いる。計測の原理は上記図27~図29の従来技術で用いられているものと同一である。

【0009】また別の従来技術として、特開平7-103734号公報に記載された「ステレオ対応探索装置」について説明する。図31は、この発明の階層的ブロックマッチング（疎密探索法）の説明図であり、図32はその発明の動作のフローチャートである。

【0010】ステレオ視覚とは、複数の異なる位置に配置された画像入力手段から入力された2つの画像の対象物上の同一点に対応する視差を計測することにより、三角測量の原理からカメラと対象点との距離を計算する手法である。図32のフローチャートのST11において画像が入力され、ST12で入力された2つの画像がそれぞれ画像メモリに格納され、ST13において、格納された画像を階層的により低い解像度に変換する処理が

行われて階層画像が生成れる。図31は第D階層の画像と第D+1階層の画像との対応関係を模式的に表したものである。この例では、第D階層の画像の1/4領域を第D+1階層の画像としている。ステップST14において、画像が小ブロックに分割され、ステップST905において、各ブロックは初期探索位置計算器によって初期探索位置が計算され、ステップST16でパターンマッチングによって他画像におけるブロックの対応点が計算され、ステップST17において、2つの画像間の対応点の位置情報を用い、視差計算器によって視差が計算される。このようにして得られた視差画像をもとに、一段階高解像度の画像を用いて再び視差が計算される。この処理を繰り返し行ない最終的には入力画像と同じ最高解像度の画像に対する対応点探索を行う。

#### 【0011】

【発明が解決しようとする課題】上記図27～図29に示された画像処理装置では、積荷の認識に必要な情報をすべて空間コード画像から獲得しており、例えば解像度256×256または512×512画素といった高い解像度の距離画像を生成する必要がある、高精度のパターン投光器が必要であり、しかもそれは時系列的に投光パターンを変化させる機構を有するものが必要であり、大規模な高価な装置となり、さらに、時系列で発生させた複数の投光パターンそれぞれに対応させて画像を入力する必要があり、大容量の画像メモリも必要とし、大規模な高価な装置となる問題点があった。またそれに要する画像入力時間の総計も多くかかるという問題点もあった。

【0012】上記図30に示す技術においても、空間コード画像のみをもとに積荷を認識するので、図27～図29の場合と同様に、大きな画像メモリも必要とし、大規模な高価な装置となる問題点があった。

【0013】さらに、図27～図29、および図30の画像処理装置は、積荷に必要な情報を全て空間コード画像から獲得するために、積荷において把持したい物体と隣接物体が密に接触している場合、すなわち、稜線の面とり量の小さい段ボール箱においては隣接する物体との境界が三次元的に明確に段差として現れないもの、あるいは、軟らかい内容物が収納されたセメント袋のような袋物の場合は、隣接する物体と密接するため、境界が三次元的には不明確であり、個々の物体を個別に分離して移送手段に把持させることが困難であるという問題点があった。

【0014】図31、図32に示された画像処理装置では、高速にステレオ対応点を探索し、距離画像を短い時間で獲得することを目的としているが、画像中の対象物の如何によらず常に画像上に画一的に設定された小ブロック単位でより粗な解像度の画像から順により密な解像度の画像へと探索を行うので、あらかじめ高解像度情報が必要な部位がわかっている場合は、逆に処理時間が長

くかかるという問題点があった。

【0015】この発明は、上記問題点を解決するためになされたものであり、大容量の画像メモリを必要とせず、簡単な装置構成で短時間で物体の三次元的な位置が把握できる画像処理装置およびこの画像処理装置を装備した物体移載装置を提供することを目的とする。

#### 【0016】

【課題を解決するための手段】この発明の請求項1に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から積載された物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報を用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものである。

【0017】この発明の請求項2に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から積載された物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報を用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えた画像処理装置と物体を把持し移載する物体移載手段とを備えたものである。

【0018】この発明の請求項3に係る画像処理装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段で構成されたものである。

【0019】この発明の請求項4に係る物体移載装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力



手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段で構成された画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0020】この発明の請求項5に係る画像処理装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されているものである。

【0021】この発明の請求項6に係る物体移載装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成された画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0022】この発明の請求項7に係る画像処理装置は、ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックのペアに関して高信頼に類似度を計算する高信頼類似度計算手段と、高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成されたものである。

【0023】この発明の請求項8に係る物体移載装置は、ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックのペアに関して高信頼に類似度を計算する高信頼類似度計算手段と、高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成された画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0024】この発明の請求項9に係る画像処理装置は、3つ以上の画像入力手段を有し、ブロック対応付手段が、3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成するものである。

【0025】この発明の請求項10に係る物体移載装置は、3つ以上の画像入力手段を有し、ブロック対応付手段が上記3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成する画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0026】この発明の請求項11に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものである。

【0027】この発明の請求項12に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段を備え、個々の対象物体の三次元位置情報を出力する画像処理装置と、物体を把持し移載する移載手段を備えたものである。

【0028】この発明の請求項13に係る画像処理装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有するものである。

【0029】この発明の請求項14に係る物体移載装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と、物体を把持し移載する移載手段とを備えたものである。

【0030】この発明の請求項15に係る画像処理装置は、物体配列の基準となるデータを格納する物体寸法データベースから自動的に生成される人工基準パターンを

用い、対象となる物体の位置を検出して、検出位置において前記対象物体に相当する画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては前記実画像基準パターンを利用するように動作するように構成したものである。

【0031】この発明の請求項16に係る物体移載装置は、物体配列の基準となるデータを格納する物体寸法データベースから自動的に生成される人工基準パターンを用い、対象となる物体の位置を検出して、検出位置において対象物体に相当する画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては実画像基準パターンを利用するように動作するように構成した画像処理装置と、物体を把持し移載する移載手段を備えたものである。

【0032】この発明の請求項17に係る画像処理装置は、段積みされた複数の対象物体の画像入力手段からの距離分布を計測する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、この第一の距離画像生成手段に格納された距離画像の情報をもとに、物体の把持位置を計算する初期把持位置計算手段と、一時的に物体を所定の位置に退避させた後に、残りの積荷の距離画像を格納する第二の距離画像格納手段と、第一の距離画像格納手段に格納された距離画像と、第二の距離画像格納手段に格納された距離画像とを比較する距離画像比較手段と、この距離画像比較手段の比較結果を基に退避した物体の把持位置を決定し、この把持位置と初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有するものである。

【0033】この発明の請求項18に係る物体移載装置は、段積みされた複数の対象物体の画像入力手段からの距離分布を計測する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、物体を把持し移動する物体移動手段と、上記距離画像の情報をもとにして物体の把持位置を計算する初期把持位置計算手段と、物体移載手段によって物体を把持位置で把持し、一時的に所定の位置に対比させた後に残りの積荷の距離画像を格納する第二の距離画像格納手段と、第一の距離画像格納手段に格納された距離画像と、第二の距離画像格納手段に格納された距離画像を比較する距離画像比較手段と、距離画像の比較結果を使用して退避した物体の把持位置を決定し初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有する画像処理装置と、物体を把持し移載する移載装置を備えたものである。

【0034】

【発明の実施の形態】

実施の形態1. 実施の形態1. の構成を示すブロック図を図1、動作の流れを示すフローチャートを図2、生成

された各段階の画像を図3に示す。ここでは、動作説明が理解され易いように認識すべき物体は、例えば段ボール箱のような多面体形状を想定し、物体は同一形状および同一サイズのものが積み重ねられている状態として説明する。

【0035】図1において、11は所定の間隔をおいて配置された2台のカメラで構成された画像入力手段、12は画像入力手段11の撮像した画像から距離画像を生成する距離画像生成手段、13は距離画像から最上段面を抽出する最上段面領域抽出手段、14は最上段物体候補抽出手段、15は原画像を格納する濃淡画像格納手段、16は物体寸法データベース、17は物体寸法データベース16に格納された情報を基に2次元パターンとしての基準パターンを自動的に生成する2次元基準パターン生成手段、18は2次元的な物体位置を検出する物体位置検出手段、19は物体位置検出手段18によって検出された物体位置と距離画像の距離情報を統合する情報統合手段である。

【0036】図3の(a)は画像入力手段11が撮像した原画像、(b)は最上段面領域抽出手段13が抽出した最上段物体領域、(c)は2次元基準パターン生成手段により物体寸法データを用いて生成された基準パターンのテンプレート、(d)は最上段物体領域から抽出された候補の一つを示す模式図、(e)はテンプレートマッチングによる法により位置決め状況を示す。

【0037】ステップST101で、装置が起動されると、画像入力手段11によって図3(a)に示す認識対象の画像が距離画像生成手段12に入力される。ステップST102では、距離画像生成手段12によって距離画像が生成される。距離画像の生成手段としては区間コード化法またはステレオ対応探索法等がある。ステップST103で最上段物体領域抽出手段13により、図3(b)に示す最上段の物体に相当する高さをもつ最上段物体領域が抽出される。ステップST104で画像入力手段11で撮影した原画像が濃淡画像格納手段15に格納され、ステップST105では2次元基準パターン生成手段17によって、物体寸法データベース16に格納された情報をもとに2次元パターンとしての基準パターンが自動的に生成される。この基準パターンはテンプレートと呼ばれ、図3(c)に示すようにテンプレートは物体の輪郭部分を表現した輪郭プレートとなっている。

【0038】ステップST106では、2値で表現された最上段物体領域の物体は画像上では矩形パターンに見えることから、直角に交わる辺縁部を検出することによって候補が抽出される。上記の通り距離画像の解像度が粗いことから、候補の位置は不正確である。ステップST107において、前記物体候補が物体位置検出手段17によって位置決めされ、位置決めには、前記2次元基準パターンが用いられ、テンプレートマッチング法による処理がなされる。ただし、大まかな位置は既に候補抽

出の段階で検出されていることから、マッチングは候補検出位置からそのごく近傍のみを探索する処理で十分である。したがってテンプレートマッチングも高速に実行可能である。位置決めの様子は図3(e)に示す。物体位置が検出されたならば、その2次元的な物体位置情報は、情報統合手段19によって前記距離画像生成手段12による距離情報と統合され、それが最終的な結果として出力される。

【0039】ステップST108では、すべての物体を検出したかどうかを判定し、もし検出していなければ処理流れのうちステップST106から再び実行され、別の物体候補について一連の処理ステップST106、ST107が実施される。もし、すべての物体が検出されていれば、ステップST109で認識処理を終了させる。以上の通り、この発明により粗く距離画像を生成し、その結果を元に物体の候補を抽出して詳細な位置検出は2次元の汎用的な処理手法であるテンプレートマッチングをそのパターン探索の範囲を非常に限定した形で実行することで、距離画像生成装置の装置規模を小さくすることができるものである。

【0040】なお、距離画像生成手段として、空間コード化法を解像度を粗くして用いるような方法、あるいは、ステレオ視覚やスリット光走査による光切断法のような方法であってもよく、粗い解像度の距離画像が生成され得る装置であればよい。

【0041】また、輪郭情報に基づいたテンプレートマッチングによる位置決めの例を示したが、同じく物体の輪郭を直線の組合せと考えると直線として辺ごとに位置合わせを行う方法を用いてもよい。また、物体寸法データベースの内容を使い、通常の2次元テンプレートマッチングを用いても良い。また、テンプレートマッチング処理内部で使用されるデータ格納形式に関して特に明示しなかったが、2次元画像として格納する方法の他に、輪郭部分のエッジの位置を1次元的に内部に羅列して格納する方法を用いてもよい。

【0042】さらに、最上段物体の候補を抽出する際に、距離画像から得られた最上段面領域情報からそのコーナー部分を利用する手法を示したが、この他にも例えば上記の2値の距離画像を用いて、別途用意した2値の距離画像としてのテンプレートを使って2値のテンプレートマッチングの方法によって粗く位置を検出する方法を適用しても同様の効果を奏する。

【0043】図2の動作フローチャートは、一例であり、図1に示した各手段の入出力関係が適正でさえあれば別の処理流れであってもよい。例えば、距離画像生成と濃淡原画像入力の手続きはどちらが先に行なわれてもよい。

【0044】以上は、認識すべき対象物体として多面体即ち箱状の物体の単一品種物体を想定した説明であったが、例えばセメント袋、米袋などのように袋状の物体で

あっても、この発明の構成を何ら変えることなく同一の構成で画像認識を実現させることができる。次に袋状物体の認識する場合の状況について、上記箱状物体における場合との違いを説明する。

【0045】図2のステップST105において、2次元基準パターン生成手段16によって2次元基準パターンが自動的に生成される。これは物体の位置決めを行うためのテンプレートマッチングにおけるテンプレート生成に相当する。袋状物体の場合は辺縁が変形するなど箱状物体とは異なり形状が不安定であるので、前記基準パターンとしては図3のような矩形の輪郭部分を取り出ただけでは不十分である。この場合は物体のコーナー部のみを輪郭で表現した部分テンプレート、あるいは物体の辺部を表現した直線輪郭部分テンプレートを自動的に生成する。いずれの場合も、袋状物体の寸法を格納した物体寸法データベース15の内容を参照して生成される。

【0046】ステップST106の物体候補抽出においては、袋状物体の場合は上で述べたような直角に交わる縁部などの情報を用いて候補を抽出することは困難である。この場合は最上段物体領域が表現された画像を2値画像とみなし、物体寸法データベース15に格納されている袋状物体モデルに関する情報から2値の物体テンプレートを自動的に生成し、これらを粗く2値テンプレートマッチングさせることで大まかな候補抽出が可能となる。

【0047】ステップST107では、物体の位置が検出されるが、ここでは前記のような物体のコーナー部分、あるいは、直線辺縁部分の輪郭テンプレートが部分テンプレートとして入力画像から得たエッジ画像とマッチングされる。物体の位置はマッチングの結果として、複数の部分テンプレートのマッチングした位置から総合的に判断、計算される、以上のように、袋状の物体に対しても箱状物体に対する認識と同じ程度に動作させることができる。

【0048】また、以上説明した実施の形態1.では、単一の品種の物体が積まれている状況を想定したが、物体寸法データベース16に複数の物体の寸法データを格納しておき、複数の2次元基準パターンを自動的に生成し、物体位置検出手段17においては前記複数の基準パターンを全てマッチングさせて最も類似性の高いテンプレートが存在する物体であるとするような動作をさせることで、異なる種類の物体が複数個積まれている状況に対しても適確に動作させることができる。

【0049】上記構成の画像処理装置を物体移載装置に組み込むことにより、高精度のパターン投光器、大容量の画像メモリを必要とせず、また画像入力時間も短時間でできる正確に動作する物体移載装置となる。

【0050】実施の形態2. 実施の形態2.の構成を示すブロック図を図4、動作の流れを示すフローチャート



を図5に示す。図6はランダムドットパターン照射の説明図、図7はステレオ画像からブロック対応付処理の方法を示す説明図である。この実施の形態2. は、実施の形態1. の距離画像生成手段の部分をランダムテクスチャパターン投光によりステレオ対応付して距離画像を生成するようにしたものであり、この点が相違し、その他の構成は実施の形態1. と同じである。

【0051】図4において、20はランダムドットパターンを投光するランダムテクスチャパターン投光手段、21は画像入力手段であり、第一画像入力手段21aと第二の画像入力手段21bとで構成されている。22はステレオ画像ブロック対応付手段である。その他の最上段画像領域抽出手段13、最上段物体候補抽出手段14、濃淡原画像格納手段15、物体寸法データベース16、2次元基準パターン生成手段17、物体位置検出手段18および情報統合手段19は実施の形態1. と同一の構成である。

【0052】以下この実施の形態2. について図5のフローチャートにしたがって動作を説明する。ステップST201で電源がONされると、ランダムテクスチャパターン投光手段20の電源が入り、認識対象物に対してランダムドットパターンが投影される。ランダムテクスチャパターン投光手段は図6(a)に示すように、パターン投光器30により、認識すべき対象物に対してランダムドットパターンを投光するものである。パターン投光器は図6のように物体の上方に下向きに設置し、図6(b)に示したようなランダムドットを認識物体に投影する。実施の形態2. では、認識すべき対象物体は、段ボール箱であるが、例えばセメント袋のような袋状の物体であっても認識可能である。対象物の上面にランダムドットが投影され、ステップST203では、第一の画像入力手段21aおよび第二の画像入力手段21bによって撮像された一对の画像のステレオ画像が入力される。このステレオ画像は、光軸を所定の距離はなして同一の解像度、同一の焦点距離の二つのカメラで撮像される。

【0053】次にランダムドットパターンについて説明する。図6(b)に示したランダムドットは、この実施\*

$$S_{SAB}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \dots\dots\dots (式1)$$

【0056】(式1)で求めた $S_{SAB}$ が最小になる( $d_x, d_y$ )を検出することで、注目ブロックbRに対応する左画像上のブロックbLが見つかったことになる。ステレオ法は、このようにして得たブロック間の位置から、その差を視差とし、三角測量の原理によってカメラから物体までの距離を得る方法である。以上説明した一連のブロック対応付処理を右画像上のすべての小ブロックに対して繰り返して行い、各ブロックごとに対応する左画像上の位置を探索し、距離画像を生成する。

\*の形態2. では、128×128の解像度で構成され、各画素は正方形であり、計算機によって一様乱数的に各画素ごとに白か黒かの値が割り当てられ、白と黒の画素数の比率はほぼ1:1となっている。パターン投光器30には投影パターンの焦点を調節する機構が設けられており、投光器と対象物体の大まかな距離がわかっている場合にはそのデータをもとに物体表面で合焦するように焦点が調節される。あらかじめ全く距離がわかっていない場合には、事前に調整された距離において焦点が合うように動作する。

【0054】ステレオ画像ペアの入力が終わると、ステップST204でランダムテクスチャの投光手段20がOFFされる。次にステップST205でステレオ画像がブロック対応付段22によってステレオ画像が処理される。図7(a)および(b)は対象物体を撮影したステレオ画像の一对の画像である。図7(a)を左画像、図7(b)を右画像と呼ぶこととする。ブロックマッチングは右画像を図のように格子状の小ブロックに分割することから行う。いま、分割されて生成された複数の小ブロックのうち、注目しているブロックbRを注目ブロックと呼ぶ、実際には左および右画像には上記ランダムドットパターンが投影されているが、図を簡略化するためにここではパターンは描いていない。この注目ブロックbRに対応する左画像上での位置を探索する。別途設定した探索範囲の中でbRと同じパターンを有するパターンを探索する。探索の方法として累積差絶対値最小化による探索を採用する。これは左画像上の探索範囲の中を注目ブロックbRを動かしながら逐次重ね合わせ、その類似性をもっとも良くなる位置をもって対応点とするものである。この際、類似性の評価としては、累積差絶対値を用いる。例えば、注目ブロックをbR( $i, j$ )、左画像をbL( $i, j$ )、ブロックサイズをN×Nとすると、左画像上の位置( $d_x, d_y$ )における累積差絶対値 $S_{SAB}(d_x, d_y)$ は、(式1)で表される。この実施の形態2. ではN=16である。

【0055】

【数1】

【0057】これ以降は実施の形態1. と同じである。ステップST206において、最上段画像領域抽出手段13によって距離画像から最上段物体面に相当する領域が抽出され、続くステップST207では、第一の画像入力手段21a、または第二の画像入力手段21bによって撮像された濃淡原画像が濃淡原画像格納手段15に格納される。この時は当然のことながらランダムドットパターンは投影されていない。ステップST208では、物体寸法データベース16に格納されている認識対象物

体の寸法データを使って、2次元基準パターン生成手段17によって2次元基準パターンが自動的に生成される。

【0058】ステップST209では、上記最上段面領域が抽出された画像は、最上段面とそれ以外の面の2つに分けられ、2値化され、最上段物体候補抽出手段14によって最上段物体候補の一つが抽出される。ステップST210では、抽出された一つの物体候補の位置近傍に関し、上記2次元基準パターンと物体位置検出手段15に格納された濃淡原画像情報を用いて、物体位置検出手段18によって候補物体の精密な位置が計測され検出される。ステップ211で、決定した候補物体の2次元位置は、情報統合手段19によって前記距離画像情報と統合され、ステップST212で全ての物体を認識したと判断するまで、ステップST209からST211までの処理が繰り返され、ステップST212の判断が全ての物体が認識したと判断されれば、最終的に物体の3次元情報として出力される。

【0059】ステップST209からST211までの処理によってすべての物体を認識したと判断されればステップST213で処理を終了する。すべての物体が認識されたかどうかの判定は、上記最上段面領域が抽出された2値化された距離画像をもとに順次物体の位置が確定したものを消去し、なおかつ画像上に残っている最上段面領域が存在するか否かを調べることによって行う。

【0060】以上説明したように、この実施の形態2.では、認識対象物体に対してランダムドットパターンを投光するので、例えば表面に図柄の存在しない無地の段ボール箱であっても、その表面各位置における距離データを得ることが可能であり、物体表面に宛先ラベルなどの位置不定なテクスチャに影響されないステレオ画像を得ることが可能である。またステレオ対応づけのためのマッチング処理も単純なブロックマッチングを適用することができる。

【0061】なお、この実施の形態2.ではカメラの解像度として、カメラ視野に対して128×128画素のランダムドットパターンを投影したが、認識対象のサイズによってはさらに細かい、または粗いドットパターンを投影しても良く、また、ブロックマッチングのブロックサイズとして16×16画素のサイズを採用したが、8×8画素、32×32画素等でもよい。また、ブロックの形状も必ずしも正方形でなくても良い。さらにマッチング時の探索範囲は、図7では長方形に設定したが、状況に応じてさまざまに変化させることでマッチングの信頼性を向上させたり処理時間を短縮することもできることは一般のステレオ対応と同様である。

【0062】さらに、この実施の形態2.では、ランダムドットパターンとして白黒の2値正方形ドットを用いたが、例えば濃淡をつけたドットパターンや着色したパターンであっても良い。また、形状についてもさまざま

な大きさのさまざまな形状のマークを組み合わせて使用しても良く、また、パターン投光器と対象物体との幾何学的位置関係については、物体に対して鉛直上方から真下に向けてパターンを投影する例を示したが、斜めから投影しても同様に距離画像が検出できる物である。

【0063】また、投光するパターンとしてランダムな位置に発生させたドットパターンを使った例を示したが、ブロック対応づけ処理において対応点を探索する場合にほとんど水平方向の探索になることを考えると、例えば周期のランダムな縦スリット群、あるいは幅のランダムな縦スリット群をパターンとして投影しても同様の効果を奏することはいうまでもない。

【0064】また、パターン投光器の構成方法に関しては、特に詳しく説明しなかったが、ランダムドットの投影位置精度は高精度が不必要であり、処理の途中で投影パターンを変化させる必要がないことから、例えば従来技術で示したように、液晶プロジェクタのような複雑な投影装置は不要であり、例えば通常の家庭用のスライド上映機のような簡単な構成の装置で十分である。このことは本発明の装置構成上の利点の一つである。

【0065】実施の形態2.は、既に述べたように、認識対象物体として段ボール箱、即ち手面体状の物体を想定して動作を説明したが、その他の形状の物体、例えばセメント袋のような袋状の物体であっても本発明の構成により容易に距離画像を生成することができ、前記本実施例における認識手順により対象物の3次元的位置を計測することができる。なお、距離画像生成より以降の処理については前述のように前記第1の発明の実施例と同様である。即ち2次元基準パターンとして物体のコーナー部分や直線輪郭部をもち、それぞれの認識対象との位置決め結果を統合することで最終的な物体位置検出が可能である。このように、本発明は物体の形状が多面体の時だけでなく、袋のような形状その他の時でも物体の3次元情報を認識することができる。

【0066】実施の形態3. 実施の形態3.の構成を示すブロック図を図8、動作の流れを示すフローチャートを図9に示す。図10は低解像度距離画像から段差領域部を抽出し、高解像度距離画像の段差領域部のみ高解像度のステレオ対応付を行う状況の説明図である。図8において、13～19は、実施の形態2.と同一機能を有するものであり説明は省略する。31は画像入力手段21が低解像度で撮像したステレオ画像から低解像度の距離画像を生成する低解像度ブロック対応付手段、32は低解像度で段差領域を抽出する段差領域抽出手段、33は低解像度の画像から抽出した段差領域部について高解像度の距離画像を生成する高解像度ブロック対応付手段、34は低解像度距離画像と、この距離画像から抽出した段差領域部の高解像度距離画像とを合成する距離画像合成手段である。

【0067】ステップST301で電源がONされ起動

すると、ステップST302で、ランダムテクスチャパターン投光手段20のランプ電源がONされ、認識対象物に対してランダムドットパターンが投影される。ステップST303では、一対のカメラで構成される画像入力手段21により一対の画像のステレオ画像が低解像度ブロック対応付手段31および高解像度ブロック対応付手段32に入力される。ステレオ画像が入力が終わるとステップST304でランダムテクスチャパターン投光手段20はOFFする。ステップST305では第一の画像入力手段21aまたは第二の画像入力手段21bから対象物体の濃淡画像が濃淡画像格納手段に15に格納される。

【0068】ステップST306では、低解像度ブロック対応づけ手段31によってステレオ画像が処理される。図10(a)は低解像度ブロック対応付手段31によって得られた低解像度距離画像部分Aを模式的に表した図である。画像入力手段として512×512画素の解像度のCCDカメラを使用しているが、低解像度距離画像としては32×32画素の距離画像を生成し、ブロックマッチングの方法は実施の形態1.で示した方法と同様の方法で行い、例えばブロックサイズを16×16画素にすることで得られる。ステップST307では上記低解像度距離画像をもとに、段差領域抽出手段32で対象物の段差領域部分Bを抽出する。段差領域部分Bの模式図を図10(b)に示す。ステップST308では、高解像度ブロック対応付手段33により、上記段差領域部分(B-Aの斜線部分)についてのみ、より高解像度の距離画像を生成する。この場合は高解像度とは128×128画素である。この処理はブロックマッチング処理においてブロックサイズを8×8に設定し、かつ、ブロックを4画素ずつずらしながら対応点を検出することで得られ、図10(c)に段差領域部分の距離画像Cを示す。図10(c)において、斜線部分(Bの枠内)は低解像度距離画像から得られた物体の段差領域、灰色の部分(Cの枠内)は、高解像度距離画像生成の結果新たに得られた距離画像の段差領域部分に相当する部分である。ステップST309では、距離画像合成手段34によって上記低解像度距離画像と高解像度距離画像を合成し、図10(d)のような128×128の解像度の距離画像が得られる。

【0069】ステップST310以降は、実施の形態2.と同様に処理される。ステップ310では、最上段面領域抽出手段13により、距離画像から最上段物体面に相当する最上段領域が抽出され、ステップST311では、物体寸法データベース16に格納されている認識対象物体の寸法データを使って、2次元基準パターン生成手段17によって2次元基準パターンが自動的に生成される。ステップST312では、上記最上段面領域が抽出された画像が、最上段面とそれ以外の面の2つに分けられて2値化され、最上段物体候補抽出手段14によ

って最上段物体候補の一つが抽出される。ステップST313では、抽出された一つの物体候補の位置近傍に関し、上記2次元基準パターンと、上記濃淡原画像情報を用いて、物体位置検出手段18によって候補物体の精密な位置が計測され検出される。このようにして決定した候補物体の2次元位置は、情報統合手段19によって上記合成された高解像度の距離画像情報と統合され、最終的に物体の3次元情報として出力する。ステップST314では、ステップST312、ST313の処理によって、すべての物体を認識したか、否かの判断を行い、たと判断されればステップST315で処理を終了となり、認識していない物体がまだ存在すると判断されれば、ステップST312、ST313に戻って物体候補を抽出する処理を繰り返す。すべての物体が認識されたかどうかの判定は、上記最上段面領域が抽出された2値化された距離画像をもとに順次物体の位置が確定したものを消去し、なおかつ画像上に残っている最上段面領域が存在するかどうか調べることによって行う。

【0070】以上のように、この実施の形態3.は、最終的に高解像度の距離画像を得るために、予め低解像度の距離画像を短時間で得ておき、物体の段差部分に相当する領域に限定して高解像度のステレオ対応付処理を行うことにして、少ない演算時間で高い解像度の距離画像を生成することを可能としたものであり、物体候補の抽出精度等も向上させることができる。

【0071】低解像度距離画像は32×32画素、高解像度距離画像は128×128画素として使用したが、許容される処理時間や認識したい対象物体の寸法などを考慮して、これと異なる解像度を設定してもよく、高速処理の低解像度距離画像生成の時間と、高解像度ステレオ対応付処理の組合せによるメリットは失われることはない。

【0072】また、低解像度距離画像生成時のステレオ画像と、高解像度距離画像生成時のステレオ画像を同じ画像として構成としたが、低解像度ステレオ画像として、フル解像度(CCDカメラ解像度と同じ解像度)の画像を1/2や1/4などに縮退させて得た縮退画像をもとにブロックマッチングを行なってもよい。さらに、高解像度距離画像生成時には第一および第二の画像入力手段21のCCDカメラのレンズをズームアップさせ、低解像度距離画像における物体の段差部分をより詳細に撮像するようにしても良い。

【0073】さらに、対象物体に投影するランダムドットパターンを一定のパターンとしたが、2種類の異なる大きさのドットパターン投光器を用意し、低解像度距離画像生成時には大きなドットパターン投影を行ない、高解像度距離画像生成時にはより小さなドットパターン投影を行うように変形して実施すれば、精度の高い距離画像が得られる。

【0074】実施の形態4. 実施の形態4. の構成を示

すブロック図を図11、動作の流れを示すフローチャートを図12に示す。図11において、13～21は実施の形態2. と同一機能を有するものであり説明は省略する。41aは第一の画像格納手段、41bは第二の画像格納手段、42は高速類似度計算手段、43は高信頼類似度計算手段、44は対応ブロック信頼度判定手段、45は距離画像格納手段である。

【0075】ステップST401で電源がONされ装置が起動されると、ステップST402でランダムテクスチャパターン投光手段20の電源がONされ認識対象物 10 に対してランダムドットパターンが投影される。ステップST403では、第一の画像入力手段21aおよび第二の画像入力手段21bによって、対象物体のステレオ画像が撮像され、第一の画像格納手段41aおよび第二の画像格納手段41bにそれぞれ入力される。第一、第二の画像入力手段21a、21bの2つのカメラは同じ解像度のCCDが使用され、同じ焦点距離のレンズが装着され、光軸は所定の距離に配置されている。画像入力後に、ステップST404でランダムテクスチャ投光手段20の電源がOFFして、物体に投影されていたランダムドット模様は消失する。ステップST405では、第一の画像入力手段41aによりランダムドットの投影されていない画像が撮像され、濃淡原画像格納手段15に格納される。この画像は後で物体位置検出に使用され\*

$$S_{SAD}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \dots\dots\dots (式2)$$

【0079】あるbRに対し、L(左)画像を探索し、 $S_{SAD}$ が最小になる位置( $d_x, d_y$ )を検出する。こうして注目ブロックbRに対応する左画像上のブロック 30 bLが検出される。対応するブロックの対が発見されると、そのペアに対し、ステップST407で、高信頼度類似度計算手段43により、上記(式2)により高信頼\*

する。

【0076】なお、第一、第二の画像入力手段21a、21bは、ステレオ視においては通常水平線上に左右に配置されており、第一の画像入力手段21aを右カメラ、第二の画像入力手段21bを左カメラと呼ぶことにする。また、それに伴い、第一の画像格納手段41aに格納された画像を右画像、第二の画像格納手段41bに格納された画像を左画像と呼ぶこととする。

【0077】ステップST406では、第一および第二の画像格納手段41a、41bに格納されたステレオ画像をもとに、高速類似度計算手段42によって高速にブロックマッチングを行う。ブロックマッチングとは、ステレオ視覚において、対応付の単位を格子状に分割した小領域(ブロック)とし、ブロック毎に左右の画像を対応付するものである。ブロックマッチングについては、実施の形態1. に記載した通りである。この際2つの画像ブロック間の類似性の評価として、累積差分絶対値を用いる。例えば、注目ブロックをbR( $i, j$ )、左画像をL( $i, j$ )、ブロックサイズをN×Nとすると、左画像上の位置( $d_x, d_y$ )における累積差分絶対値  $S_{SAD}(d_x, d_y)$  は、(式2)で表される。

【0078】

【数2】

※の類似度が計算される。この時、左画像は走査されない。(式2)を評価して得た( $d_x, d_y$ )に対し、次の(式3)により類似度を評価する。 $S_{corr}$ は一般に正規化相互相関関数と呼ばれるものである。

【0080】

【数3】

$$S_{\text{CORR}} = \frac{C}{A \cdot B}$$

24

(式3)

$$A = \sqrt{N^2 \sum_{j=1}^N \sum_{i=1}^N \{L(d_x + i, d_y + j)\}^2 - \left\{ \sum_{j=1}^N \sum_{i=1}^N L(d_x + i, d_y + j) \right\}^2} \quad \text{.. (式4)}$$

$$B = \sqrt{N^2 \sum_{j=1}^N \sum_{i=1}^N \{b_R(i, j)\}^2 - \left\{ \sum_{j=1}^N \sum_{i=1}^N b_R(i, j) \right\}^2} \quad \text{..... (式5)}$$

$$C = N^2 \sum_{j=1}^N \sum_{i=1}^N \{L(d_x + i, d_y + j) \cdot b_R(i, j)\} \\ - \sum_{j=1}^N \sum_{i=1}^N L(d_x + i, d_y + j) \sum_{j=1}^N \sum_{i=1}^N b_R(i, j) \quad \text{..... (式6)}$$

【0081】(式3)は、(式2)に比べより正確で信頼性の高い類似度を計算するものである。得られた類似度 $S_{\text{CORR}}$ は、上記対応付結果から得られた視差、即ち距離情報とともに距離画像格納手段45に格納される。

【0082】ステップST408では、対応ブロック信頼度判定手段44により、上記高信頼な類似度が予め設定され計算機内部に格納されていた類似度しきい値と比較され類似度が十分大きい場合には正しい対応付がなされたと判定され、類似度がしきい値よりも小さい場合には誤った対応付であったと判定される。判定結果は距離画像格納手段45に格納される。続くステップST409で、右画像のすべてのブロックについて高速な類似度と高信頼の類似度を用いた対応付を行ない対応付の信頼度を判定していない場合には、再び処理ステップST406より同様の処理を繰り返す。

【0083】ステップST410以降は、実施の形態2.と同様に処理される。ステップ410では、最上段面領域抽出手段13により、距離画像から最上段物体面に相当する最上段領域が抽出され、ステップST411では、物体寸法データベース16に格納されている認識対象物体の寸法データを使って、2次元基準パターン生成手段17によって2次元基準パターンが自動的に生成される。ステップST412では、上記最上段面領域が抽出された画像が、最上段面とそれ以外の面の2つに分けられて2値化され、最上段物体候補抽出手段14によって最上段物体候補の一つが抽出される。ステップST413では、抽出された一つの物体候補の位置近傍に関し、上記2次元基準パターンと、上記濃淡原画像情報を用いて、物体位置検出手段18によって候補物体の精密な位置が計測され検出される。このようにして決定した候補物体の2次元位置は、情報統合手段19によって上

記合成された高解像度の距離画像情報と統合され、最終的に物体の3次元情報として出力する。ステップST414では、ステップST412、ST413の処理によって、すべての物体を認識したか、否かの判断を行いしと判断されればステップST415で処理が終了となり、認識していない物体がまだ存在すると判断されれば、ステップST412、ST413に戻って物体候補を抽出する処理を繰り返す。

【0084】以上ように、この実施の形態4.では、ステレオ画像をブロック単位で対応付する際に、まず最初に高速な類似度評価手段を使って画像探索を行い、求まった対応ブロックの対について一度だけ高信頼な類似度を評価する手段によって対応ブロックの類似性を精密に求めることによって、時間のかかる探索処理を高速に行い、かつ、最終的に対応ブロックが正しいかどうかの判定は高信頼で行うことができる。

【0085】なお、この実施の形態4.では、高速な類似度計算式として累積差分絶対値を用い、高信頼の類似度計算式として正規化相互相関を用いたが、例えば、高速な類似度としてブロック間の各画素値の差の絶対値の最大値を使用したり、高信頼類似度として、差の自乗の総和を使用してもよい。実施の形態4.の本質は、信頼性よりも高速性を重視した類似度計算手段と、その逆の特性を持つ類似度計算手段との組合せによって効果を奏するものである。

【0086】また、この実施の形態4.は、高速な類似度計算と、高信頼な類似度計算において、同じように画像ブロックの各画素値を使用したが、高速類似度計算時には使用する画素をブロックから間引き操作によって1/2または1/4といったように限定して選択してもよく、その場合にはたとえ上記2種類の類似度計算に使用

される類似度評価式が同一のものであっても、高速類似度計算が高速に実行されるので同様の効果を奏する。

【0087】さらに、この実施の形態4.は、すべての処理を計算機によって実行することを前提と、利用可能な特別のハードウェアによってある種の類似度計算が高速に実行せられる環境が提供された場合には、高速類似度計算にはそのハードウェアを用いて高信頼類似度計算時には別の高信頼性が確保される方法で類似度を計算しても良い。

【0088】実施の形態5. 実施の形態5.の構成を示すブロック図を図13、動作の流れを示すフローチャートを図14に示す。図13において、13~20は実施の形態2.と同一機能を有するものであり説明は省略する。51は画像入力手段であり、第一の画像入力手段51a、第二の画像入力手段51b、第三の画像入力手段51cの3台備えている。52はステレオ画像ブロック対応付手段である。

【0089】ステップST501で電源がONされ装置が起動すると、ステップST502でランダムテクスチャパターン投光手段がONされ、認識対象物に対してランダムドットパターンが投影される、ステップST503~ST505では、第一の画像入力手段51a、第二の画像入力手段51b、および第三の画像入力手段51cによって順次3枚の画像が入力される。3つの画像入力手段51a、51b、51cはCCDカメラであり、これら3つのカメラは同じ解像度で、同じ焦点距離のレンズが装着された光学的仕様は同じものが使用される。カメラの配置は、それぞれのカメラの光軸が空間上の同一平面上にあり、かつどの光軸も交わらないように平行に配置されている。ここでは便宜上3つのカメラはそれぞれ左カメラ、中央カメラ、右カメラと呼ぶことにする。光軸間の距離は、左カメラと中央カメラの間隔は中央カメラと右カメラの間隔に等しく配置する。このカメラ配置は、上記実施の形態2.~4.の場合の2つのステレオ配置カメラの中央に3つめのカメラを配したことに相当する。

【0090】画像入力が完了すると同時に、画像はそれぞれ別の画像格納手段である画像格納手段に格納され、ステップST507でランダムテクスチャパターン投光器20のランプ電源がOFFし、物体表面のランダムドットパターンは消える。ステップ407では、中央の第二の画像入力手段51bによって、対象物の濃淡画像が撮像され、濃淡原画像格納手段15に格納される。

【0091】ステップST508では、ステレオ画像ブロック対応付手段52によってステレオ画像の対応付処理を行う。この処理の流れについては、図15に詳細な動作のフローチャートを示す。ステレオ対応づけ処理がST508-1でスタートして最初のステップST508-2では、第一の画像入力手段51a（左カメラ）による画像と第二の画像入力手段51b（中央カメラ）に

よる画像との間で対応点の探索を行う。この処理は2つの画像のブロック対応付処理であり、上記実施の形態2で示した処理と同様に行う。この実施の形態5.では、中央カメラの撮像画像をブロック小領域に分割し、各ブロックごとに左カメラ上で対応点を探索する。図16はこの2つの画像間のブロック対応付の様子を模式的に示したものである。図16(a)は、認識対象物体としての段ボール箱が映っている画像を示している。梱包されたダンボール箱の場合、表面に透明のフィルム上のものが貼り付けられることがしばしばある。その場合、パターン投光器20や室内の他の照明装置からの光、特に乱反射光によって、図16(b)のように情報欠落と判定されるブロックが発生することがある。この情報欠落部とは、例えば中央カメラ51bではテーブル部に乱反射が起こっていないのに、左カメラ51aでは起こっている状態があり、当該ブロックに対応する画像パターンが他方の画像の中には発見できないという事実が発生する。この状態における処理はステップST508-3において情報欠落ブロックか否かを判定することで実行される。

【0092】ステップST508-4では、中央カメラが撮像した第二の画像と、右カメラが撮像した第三の画像との間でブロック対応付が行なわれる。中央カメラによる画像をブロックに分割し、各ブロックごとに対応する点を右カメラ画像上で探索する。この時も上記の対応点探索と同様に情報欠落ブロックが発生する。さらに、ステップST508-5では、上記中央カメラ51bと右側カメラ51cとの組合せで得られた情報欠落点に関する情報が、上記の左カメラ51aと中央カメラ51bとの情報欠落データが組み合わされる。この時、左-中央ペアにおける情報欠落ブロックの位置と、中央-右ペアにおける情報欠落ブロックの位置との間には何ら直接的な関係はない。どちらのペアにおいても、中央カメラ51bの画像を基準として、左または右カメラ画像上を探索しているので、各ペアの情報欠落ブロック情報は、中央カメラ画像上のブロック単位で与えられ、容易に比較できる。

【0093】いま、左-中央ペアにおいて情報が欠落していると判定されたブロックが、中央-右ペアにおいては情報欠落でないと判定されたならば、当該ブロックは情報欠落でないとし、その視差情報としては情報欠落でないペアを用いて計算される。両ペアともに情報欠落と判定された場合には当該ブロックは情報が欠落していると判定し、視差情報は格納しない。中央カメラの全ブロックに対して同様の処理が行なわれ、ステップST508-6でステレオ対応づけ処理が終了する。即ち、左-中央ペア、および中央-右ペアの両方において情報欠落ブロックになっていない限り、視差情報、即ち距離情報を得ることが可能である。

【0094】なお、ステレオ対応探索における類似点探



索において用いられる類似度計算手段としては、次の  
(式7)の累積差分絶対値 $S_{SAB}(d_x, d_y)$ を用い  
る。

$$S_{SAB}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \dots\dots\dots (式7)$$

【0096】次に、図14のステップST509では、  
上記ステレオブロック対応付処理において生成された距  
離画像を用いて、最上段面領域抽出手段13によって物  
体の最上段面の領域を抽出する。

【0097】ステップST510以降は、実施の形態  
2.と同様に処理される。ステップST510では、物  
体寸法データベース16に格納されている認識対象物体  
の寸法データを使って、2次元基準パターン生成手段1  
7によって2次元基準パターンが自動的に生成される。  
ステップST511では、上記最上段面領域が抽出され  
た画像が、最上段面とそれ以外の面の2つに分けられて  
2値化され、最上段物体候補抽出手段14によって最上  
段物体候補の一つが抽出される。ステップST512では、  
抽出された一つの物体候補の位置近傍に関し、上記  
2次元基準パターンと、上記濃淡原画像情報を用いて、  
物体位置検出手段18によって候補物体の精密な位置が  
計測され検出される。このようにして決定した候補物体  
の2次元位置は、情報統合手段19によって上記合成さ  
れた高解像度の距離画像情報と統合され、最終的に物体  
の3次元情報として出力する。ステップST513では、  
ステップST511、ST512の処理によって、  
すべての物体を認識したか、否かの判断を行い、したと  
判断されればステップST514で処理を終了となり、  
認識していない物体がまだ存在すると判断されれば、ス  
テップST511、ST512に戻って物体候補を抽出  
する処理を繰り返す。

【0098】以上説明したように、この実施の形態5.  
では、画像入力手段51のカメラを3台使って2組のス  
テレオ画像を生成し、一方の組のステレオ画像で生じた  
情報欠落ブロックに対して、他の組のステレオ画像画像  
による視差情報によって、情報を補完して、情報の欠落  
が防止される。

【0099】なお、この実施の形態5.では、3つの画  
像入力手段の空間的配置として、同一平面に等間隔に平  
行して存在する光軸を有するように配置したが、左ある  
いは右、あるいはその両方のカメラ光軸を中央カメラ光  
軸と交わるように配置したり、カメラ間隔を等間隔でな  
いように配置したり、または3つの光軸が必ずしも同一  
平面上にないように配置しても、同様に欠落ブロックの  
補完する効果は得られる。

【0100】なお、実施の形態5.では、高速な類似度  
計算式として累積差分絶対値を用い、高信頼の類似計算

\* 【0095】  
【数4】

式として正規化相互相関を用いたが、例えば、高速な類  
似度としてブロック間の各画素値の差の絶対値の最大値  
を使用したり、高信頼類似度として、差の自乗の総和を  
使用してもよい。

【0101】実施の形態6. 実施の形態6. の構成を示  
すブロック図を図17、動作の流れを示すフローチャー  
トを図18に示す。図17において、11~13、15  
~17、19、21は実施の形態1. と同一機能を有す  
るものであり説明は省略する。図において、64は物体  
候補組合せ仮説生成手段、65は仮説妥当性検証手段、  
69は物体位置検出手段である。

【0102】ステップST601で、装置が起動され  
ると、画像入力手段11の2つのカメラで対象物体の画像  
が距離画像生成手段12に入力され、個別に格納され  
る。2つのカメラは、光軸を共有しない同一の仕様のも  
のが使用される。ステップST602で、距離画像生成  
手段12によって上記2枚のステレオ画像がステレオマ  
ッチングされ、その視差から距離画像が生成される。

【0103】ステップST603で、画像入力手段11  
により撮像された濃淡原画像が濃淡原画像格納手段に格  
納される。この実施の形態6. の認識対象とする物体の  
濃淡原画像を図19(a)に示す。物体は段ボール箱等  
の箱状の物体であり、それらが何段かに積載されてい  
る。ステップST604で距離画像はから、最上段面領  
域抽出手段13によって、最上段物体の上面部分の画像  
から領域として抽出される。図19(b)は、抽出され  
た最上段物体領域である。図のように、最上段以外の領  
域も示されている。

【0104】ステップST605では、箱状物体の寸法  
データを格納した物体寸法データベース16に格納され  
たデータを用いて、2次元基準パターン生成手段17に  
よって物体の2次元的な画像パターンを表現した基準パ  
ターンを自動的に生成する。図19(c)にその基準パ  
ターンの模式図を示す。

【0105】以下の動作の説明が、この実施の形態6.  
の本質となる部分である。ステップST607におい  
て、物体候補組合せ仮説生成手段64によって、上記抽出  
された最上段面領域と2次元基準パターンとから、物体  
候補の組合せ仮説が一つ生成される。仮説とは図19  
(c)に示されたようなものであり、この実施の形態  
6. では、この仮説を「仮説1」と呼ぶ。仮説は図19  
(b)のような最上段物体領域を画像で表現したものに

対し、認識対象物の基準パターンをパターンマッチング、例えばテンプレートマッチングの技法によって処理し、図19(b)のような最上段面領域を形成する2次元基準パターンの組合せを類推することによって得られる。また、同時に、物体位置検出手段69によって図19(d)のように物体2次元的位置が検出される。物体位置検出は、上記格納された濃淡原画像と上記2次元基準パターンとの比較によって行う。2次元基準パターンの輪郭部分の情報をを用い、上記濃淡画像から検出したエッジ画像との間でテンプレートマッチングを行う。その際、仮説生成の段階で物体の粗い位置は既知なので、テンプレートマッチング処理におけるパターン探索領域は上記位置近傍のみでよい。

【0106】ステップST608では、上記仮説が妥当であるか否かが検証される。仮説妥当性検証手段65は図19(d)のような仮説画像を生成し、物体候補1から6までの6個の物体の組合せによって、図19(b)のような最上段面領域が合理的に生成され得るかどうかを計算する。即ち、上記6個の物体候補によって生成される仮説画像と、最上段面画像とを比較し、その差となる領域の面積を算出して上記面積が大きいほど誤った仮説であると判定するように動作する。ステップST608ではこのようにステップST606で生成した一つの仮説に対し、その妥当性指標を一つ与える。上記面積の逆数が妥当性指標となる。

【0107】ステップST609では、ステップST606からST608までの処理を考えられる仮説をすべて生成するまで繰り返し、ステップST610では、仮説をすべて生成し、それらの中でもっとも妥当性の高い仮説を選択する。即ち、各仮説についての前記妥当性指標を比較し、最も高い指標を持つ仮説を選ぶ。図19

(e)(f)は、それぞれ別の仮説2および仮説3である。図から明らかなように、この場合は仮説2が最も高い妥当性指標を持つのでこれが最もよい仮説として、ステップST610において選択されることになる。

【0108】ステップST611では、情報統合手段19によって、上記の最もよい仮説に相当する2次元位置情報と前記距離画像が統合され、ステップST612で認識結果として装置から出力されて処理を終了する。

【0109】この実施の形態6.は、距離画像生成手段として2つのカメラによるステレオ視を用いたが、3つ以上のカメラによるステレオ視、ランダムドット等の特徴ある投光パターンの投光を併用したステレオ視、あるいは従来技術として説明した空間コード化法による距離画像生成手段でもよい。この実施の形態6.は、距離画像から得た認識結果の仮説を検証して尤もらしい結果を得ることに特徴があるので、距離画像の生成方法はどんなものでもよい。

【0110】また、この実施の形態6.は、仮説の妥当性を検証する方法として、仮説画像と実際の画像との差

をもとに妥当性指標を計算する方法を用いたが、例えば、上記仮説を得る際に計算されたテンプレートマッチングの類似度のすべての候補物体に対する総和が大きいほど妥当性が高いとするような指標を用いても同様の効果を得ることができる。

【0111】また、この実施の形態6.では物体位置の検出方法として、基準パターンから得た物体の輪郭情報と、濃淡原画像から生成したエッジ画像とをテンプレートマッチング手法により処理していたが、これを2値画像のテンプレートマッチングなどの位置決め手法を用いても同様の効果を得ることができる。

【0112】さらに、この実施の形態6.では、物体候補の組合せ仮説の生成と検証を逐次行うような処理フローを説明したが、多数の仮説を一度に生成して格納しておき、それらの妥当性検証を一度に行なって最も妥当性の高いものを選択するようにしてもよい。

【0113】さらに、実施の形態6.では、すべての仮説を生成するように説明したが、仮説は物体候補の組合せとして得られることから、物体の組合せ最適化問題として扱い、近似解を求めても同様の効果が得られる。

【0114】実施の形態7. 実施の形態7.の構成を示すブロック図を図20、動作の流れを示すフローチャートを図21に示す。図20において、11~13、15~19は実施の形態6.と同一機能を有するものであり説明は省略する。71は物体配列基準データベース、72は物体配列判定手段、73は警告発生手段である。

【0115】この実施の形態7.は、実施の形態6.の構成に、物体配列が誤って配列されているときに警告を発生するように構成したものである。したがって、図21のフローのステップST701~ST711は、図18のステップ601~611と同じである。

【0116】ステップST701で装置が起動されると、画像入力手段11によって対象物体の画像が距離画像生成手段に入力され、個別に格納される。ステップST702で、距離画像生成手段12によって上記2枚の画像がステレオマッチングされ、その視差から距離画像が生成される。この処理は、必ずしも単純なステレオマッチングである必要性はなく、例えば実施の形態2.に示したようなパターン投光器を併用したステレオ視覚であつてもよい。

【0117】ステップST703では、画像入力手段11により認識対象物の濃淡原画像が入力され、濃淡原画像格納手段15に格納される。ステップST704で最上段面領域抽出手段13により最上段面領域が抽出される。ステップST705では、箱状物体の寸法データを格納した物体寸法データベース16を用いて2次元基準パターン生成手段17によって物体の2次元的な画像パターンを表現した基準パターンを自動的に生成する。ステップST706では、物体候補組合せ仮説生成手段64によって、上記抽出された最上段面領域と基準パター



ンとから、物体候補の組合せ仮説が一つ生成される。仮説とは実施の形態6. で説明したと同様の図19(d)に示されたようなものである。この実施の形態7. においては、仮説は図19(b)のような最上段物体領域を画像で表現したものに対し、認識対象物の基準パターンをパターンマッチング、例えばテンプレートマッチングの技法によって処理し、図19(b)のような最上段面領域を形成する2次元基準パターンの組合せを類推することによって得ている。ステップST707では、物体位置検出手段69によって図19(d)のように上記仮説に含まれる候補物体それぞれの2次元的位置が検出される。物体位置検出は、上記格納された濃淡原画像と、上記2次元基準パターンの輪郭部分の情報をを用い、濃淡画像から検出したエッジ画像との間でテンプレートマッチングによって検出される。

【0118】ステップST708では、上記複数の物体候補によって生成される仮説画像と、最上段面画像とを比較し、その差となる領域の面積を算出して大きいほど誤った仮説である可能性が高いと判定する。ステップ709では、ステップST706~708の処理を考えられる仮説をすべて生成するまで繰り返し、ステップ710では、仮説をすべて生成し、それらの中でもっとも妥当性の高い仮説を選択する。即ち、各仮説についての前記妥当性指標を比較し、最も高い指標を持つ仮説を選ぶ。

【0119】ステップST711では、情報統合手段19によって上記仮説で最もよい仮説に相当する2次元位置情報と上記距離画像が統合される。この時点で、この実施の形態7. における画像処理装置は物体の配列および各々の物体の3次元的位置を知ったことになり装置内部に格納される。ステップ712では、物体配列判定手段72によって上記格納された認識結果が、予め用意された正しい配列データと比べて合致しているか否かが判定される。正しい配列データは、物体配列基準データベース71に格納された被認識物体によって決められた配列パターンである。例えば、この実施の形態7. による装置が使用されるような物流工場においては、認識対象物体としての段ボール箱はその形状や寸法によって予め正しい「積みパターン」と呼ばれるパレット上への箱の積み方が決められている。

【0120】上記物体配列基準データベースにはそのような基準とすべき物体の配列が記述されている。ステップST712において認識した結果が基準となる積みパターンと合致していない場合には、ステップST713において、警告発生手段73によって装置使用者や上位システムの管理者に対して警告を発生して物体が正しく積まれていないことを知らせる。この実施の形態7. は、ブザーによる音と、装置に接続された表示用ディスプレイ装置への文字表示によって警告が発せられる。ステップST712において認識結果が正しいパターン

であると判定された場合は、情報統合手段19によって最終的な物体認識結果が出力され、例えばロボットにデータが送信されることによりロボットによる物体の把持動作が行なわれることになる。認識動作としてはステップST715で動作を終了する。

【0121】この実施の形態7. では、このように物体を認識するだけでなく、その配列パターンを基準データと比較することによって予想された配列なのかどうかを判定し、さらに上記認識結果が基準データと異なる場合には警告を発生する機能を有している。これにより、作業者が自分の積み作業ミスに気づいたり輸送中の激しい荷崩れを検知できるという利点がある。

【0122】なお、距離画像生成手段として2つのカメラによるステレオ視を、仮説の妥当性検証の方法として仮説画像と実際の画像との差をもとに検証する方法を、そして物体位置の検出方法として物体の輪郭情報を利用したテンプレートマッチング手法をそれぞれ利用している。しかし、これら部分的な処理手段処理手段は、他の手法により実現しても同様の効果を奏することはいうまでもない。例えば、距離画像生成手段として、実施の形態2. で説明したパターンを併用するステレオ視を、仮説の妥当性検証方法として各物体を濃淡基準パターンと比較した類似度の総和が高いほど妥当性が高いと判断されるような方法を物体位置検出方法として仮説に含まれる各物体を予め用意した濃淡画像パターンと正規化相互相関によるテンプレートマッチング法を、それぞれ採用してもよい。

【0123】また、認識結果が予め格納された物体配列基準データベースと比較して合わないと判断した時に警告を発生する例を説明したが、基準データとの類似度合に応じて、警告の内容を変えて装置使用者や作業者に知らしめるようにしてもよい。例えば、制御表示器に『自動認識の結果判定された積み方の正しさはレベル8です』などと表示する。さらに認識した結果、積荷の荷崩れ状態が非常に顕著であり、このまま作業を継続することが適当でないと判断される場合には上記警告を発するだけでなく、上位の管理計算機にシステムの停止を要求したり、または本装置が直接システム停止したりするようにしてもよい。

【0124】さらに、警告手段としてブザーと表示器による文字表示を採用したが、この他にも人間に警告を送る手段として、合成された人間の声を発生させるなどの方法を用いてもよい。

【0125】実施の形態8. 実施の形態8. の構成を示すブロック図を図22、動作の流れを示すフローチャートを図23に示す。図22において、11~15、17、19は実施の形態1. と同一機能を有するものであり説明は省略する。図において、81は物体の寸法データをもとに人工2次元基準パターンを生成する人工基準パターン生成手段、82は濃淡原画像から認識対象物体

の寸法に対応する実画像の基準パターンを抽出する実画像基準パターン切り出し手段、83は実画像基準パターン格納手段、88は物体位置検出手段である。

【0126】ステップST801で装置が起動されると、画像入力手段11によって認識対象物体の画像が距離画像生成手段12に入力される。ステップST802では、入力された画像を用いて、距離画像生成手段12によって距離画像が生成される。距離画像生成手段としてはどのようなものでもよいが、この実施の形態8.では2つのカメラによるステレオ視覚を利用することとする。したがって、画像入力手段は2つのテレビカメラを用いる。これら2つのカメラは同一の光学的仕様を持ち、光軸を共有しないように配置される。通常は人間の目のようにカメラ光軸がほぼ平行になるように配置する。ステップST803では、上記画像入力手段11の一方のカメラを使用し、濃淡原画像を撮像して濃淡原画像格納手段15に格納する。この原画像は、上記距離画像生成時に撮像した濃淡画像をそのまま使用しても差し支えない。

【0127】ステップST804では、距離画像を用いて最上段面領域抽出手段13により物体の最上段面領域を抽出する。抽出した最上段物体の領域は、図3(b)に模式的に描かれているものと同様である。つまり、最上段領域は「1」、それ以外の領域は「0」というように2値の画像として表現される。なお、この実施の形態8.においては認識対象を段ボール箱のような箱状の物体であるとし、地上に山積みされた同一の物体を天井に下向きに設置された画像入力手段、即ちテレビカメラで撮像している。ステップST805では、物体寸法データベース17に予め格納されている認識対象となる箱状の物体の寸法データをもとに、人工基準パターン生成手段81によって人工2次元基準パターンが生成される。これは実施の形態1.の説明における2次元基準パターンと同様のものであるが、この実施の形態8.では特にこれが人工的、自動的に生成されたパターンであることを強調するためにこのように呼ぶ。この基準パターンは図3(b)に示されているような物体の輪郭部分のみを表現したものである。

【0128】ステップST806では、抽出された最上段面領域を表現した2値画像から最上段物体候補抽出手段14によって最上段に存在すると考えられる物体の候補が一つ抽出される。ここでは、2値で表現された最上段物体領域に対し、物体が画像上では矩形パターンに見えることから、直角に交わる辺縁部を検出することによって候補を抽出する。抽出された候補の模式図は実施の形態1.の説明における図3(d)と同様である。ステップST807において、上記物体候補が、物体位置検出手段14によってテンプレートマッチング法を用いることで位置決めされる。テンプレートとしては、上記人工基準パターンが用いられる。また、テンプレートマッ

チングにおいて、テンプレートを探す対象となる画像としては、上記濃淡原画像格納手段15に格納された原画像が物体位置検出手段88内においてエッジ検出処理された結果得られるエッジ画像が用いられる。このテンプレートマッチングにおいては、大まかな位置は既により上記候補抽出の段階で検出されていることから、ここでのマッチングは候補検出位置からそのごく近傍のみを探索するだけの処理で十分である。位置決めの様子は図3の通りである。

【0129】位置決め処理の結果としては、位置情報とともにテンプレートとの類似度が得られるが、これは予め生成された人工基準パターンとの類似性を表現したデータである。類似度が高いほど検出された物体は予め想定した正しい箱に似ており、類似度が低い場合には位置決めに失敗している可能性が高いといえる。そこで、ステップST808では、得られた類似度データを物体検出の信頼度であると見做し、これが予め本装置内に格納されている信頼度しきい値より大きいかなんかを判断する。位置決めの結果得られた物体の信頼度が、信頼度しきい値より小さければ誤った物体が検出されたと判断し、ステップST806から処理を繰り返し、再び別の最上段物体候補を抽出して人工パターンを使った位置検出を行う。得られた信頼度がしきい値より大きい場合は次のステップST809に進むことになる。

【0130】ステップST809では、実画像基準パターン切り出し手段82によって、最終的に検出された物体位置における認識対象物体の寸法に対応する実画像の基準パターンが抽出され、実画像基準パターン格納手段83格納される。この時、実画像とは、濃淡原画像格納手段15に格納された濃淡原画像を意味し、切り出し位置は上記検出された物体の位置である。また、切り出し領域は矩形であり、その寸法は上記物体の実際の寸法データを上記距離画像から得られたカメラと物体との距離によって画像上の寸法に変換したものである。即ち、切り出されるのは認識したい物体1個に対応する濃淡画像としての矩形画像領域である。

【0131】ステップST810では、物体位置検出手段88によって、画像から2つめ以降の物体が検出される。ここでは、前記一つめの物体検出時と異なる手法が用いられる。上記実画像基準パターン格納手段83に格納された濃淡実画像としての物体ひとつに相当するパターンをテンプレートとし、濃淡原画像格納手段83に格納された原画像中からテンプレートに最も似ているパターン位置を検出するというテンプレートマッチング法を適用する。

【0132】ステップST811においては、検出された物体が予め認識したいと要求された全ての物体に相当するかなんかが判定され、全ての物体が認識されたならば情報統合手段19によって上記検出物体位置と距離画像情報を統合して各物体の3次元位置を計算し、上位ホス

トコンピュータまたはロボット等にその結果を送信した後、ステップST812で処理を終了する。物体の認識が終わるまでステップST810から処理が繰り返される。即ち、ひとつめの物体は距離画像等を使って認識されるが、2つめ以降の物体は1つ目の認識時に得た濃淡画像としての実画像パターンをもとにテンプレートマッチングで検出する。

【0133】なお、実施の形態8.では、距離画像生成手段として一般的なステレオ視覚を採用した例で説明したが、実施の形態2.で説明したようなランダムドットパターン投光を用いたステレオ視や、従来の技術で説明した空間コード化法による距離画像生成でも同様の効果は得られる。

【0134】実施の形態9. 実施の形態9.の構成を示すブロック図を図24、動作の流れを示すフローチャートを図25に示す。図24において、11は画像入力手段、91は距離画像生成手段、92aは第一の距離画像格納手段、92bは第二の距離画像格納手段、93は距離画像比較手段、94は初期把持位置計算手段、95は位置補正量計算手段、96は物体移動手段である。

【0135】物体の形状は、説明が理解され易いように例えば段ボール箱のような多面体で、同一形状、同一寸法のものを想定して説明する。この実施の形態9.は、認識対象物体の形状が必ずしも多面体形状である必要はなく対象がセメント袋等のような袋状の物体であってもこの実施の形態9.の構成を使用することができる。

【0136】ステップST901で装置が起動されると、画像入力手段11によって認識対象の物体の画像が距離画像生成手段91に入力される。ステップST902では、入力された画像を用いて、距離画像生成手段91により距離画像が生成され、ステップST903において第一の距離画像が生成され、第一の距離画像格納手段92aに格納される。距離画像の生成手段としてはどのようなものでもよいが、例えば空間コード化法や、ステレオ視覚から生成する方法がある。

【0137】距離画像について図26によって説明する。図26(a)には初期状態、即ちステップST901が実行されたときの箱状の物体が段積みされた状態の、画像入力手段11で撮像された原画像が模式的に表現されている。図26(b)は第一の距離画像格納手段92aに格納された距離画像を模式的に表したものである。距離画像は、画像入力手段のカメラから物体までの距離を画像的に表したものであり、カメラからの距離が近い程明るい色で表し、カメラからの距離が大きくなるほど暗い色調で表現している。

【0138】次のステップST904では、初期把持位置計算手段94によって距離画像から初期把持位置が計算される。初期把持位置とは、ロボット等の物体移動手段96によって物体を移動するために把持する位置である。初期把持位置を決定するには上記第一の距離画像を

処理し、最上段に位置する物体領域即ち図26(b)において明るい色で表示されている領域からコーナ部分を認識しコーナ位置から一定オフセットをかけた位置を初期把持位置とする。この場合は、物体A上に設定されたことになる。図26(b)上で矢印の方向は把持方向、矢印と直線の交点が把持位置を表している。ただし、ここで意図的に物体Aを把持すべく把持位置を決定したわけではなく、距離画像から単純な方法で求めた把持位置がこの場合は物体A以外の物体上に存在したとしても、結果として初期把持位置がどれか一つ物体上に設定されたのであれば、積み荷の移載作業において何ら問題ではない。

【0139】ステップ905では、初期把持位置を、物体異動手段たるロボットが把持し、画像入力手段の視野外に物体を一時的に退避する。図26(c)は物体退避後の様子を示している。物体の退避動作は、ロボットが物体A上の把持位置を、吸着ハンドで吸着し、予め決められた退避位置までハンドを移動させることにより行う。ただし、ここでは、ロボットは物体Aを把持したままハンドを静止させることで退避させている。予め決められた退避位置とは、ハンドがその物体を静止させている位置が、上記画像入力手段11の視野の外にあると明確に解っている位置であり、また、そのように設定される位置である。

【0140】ステップST906では、物体退避後に再び距離画像生成手段91によって積荷の距離画像が生成され、ステップ907で、第二の距離画像格納手段92bに格納する。図26(d)はこの実施の形態9.の第二の距離画像を模式的に表現したものである。この図では、最上段物体領域から、退避された物体Aに相当する領域がかけている。

【0141】ステップST908では、距離画像比較手段93により、第一の距離画像と第二の距離画像が比較され、さらにその結果をもとに位置補正量計算手段95により位置補正量を計算する。この実施の形態9.では、距離画像の比較として単純な差分を行う。第一の距離画像から第二の距離画像を引くことで図26(e)が検出される。この図は図26(c)一時退避した物体Aに相当する領域Bのみが2つの距離画像の差として残ったものである。ロボットによって把持され退避された物体は一つだけであり、退避動作の前後では、距離画像において、退避された物体に相当する部分のみが差として存在するのは当然である。距離画像の差分結果として領域Bは物体一個に相当することがわかっているので、簡単にその形状や重心位置を計算することができる。画像上で明確に背景と分離された領域の重心や形状、向きを計算する手法としては、例えば差分画像を2値化して領域毎にラベリングを施した後各ラベル付けされた領域毎に重心やモーメント最大方向を計算する従来からの一般的に使用されている手法が適用できる。

【0142】図26(f)は、退避した物体の拡大図であるが、初期把持位置の他に、距離画像差分結果から得られた新しく計算した把持位置が求まっており、上記重心位置を新しく計算した把持位置としている。位置補正量はこれら2つの把持位置の差として計算される。

【0143】この時点において、ロボットはいまだ初期把持位置を把持したままであるが、正しい把持位置すなわち物体の重心位置との差は既に計算されている。したがって、上記位置補正量データをロボットに送ることによりロボットはステップST909において、退避した物体Aについて、ハンドが物体のどの位置を把持しているかを知ることになる。ゆえに、続くロボットによる移動動作の際には、ロボットは上記補正量を考慮した移動動作を行うことができる。これは、積荷の移動作業において、ロボットにとってはあたかも物体を最初に把持した時点で既に物体の重心位置即ち把持すべき位置を正確に知っていたことに相当し、ロボットが物体の正確な把持位置を把持して移動作業を実行することと同様の動作を実現していたことになる。

【0144】ステップST910で全ての物体を認識したかどうか判断し、もし認識したならステップST911に進んで処理を修了する。処理が未終了の場合は処理が終了するまでステップST902から処理を繰り返す。

【0145】以上のように、この実施に形態9.は、第一の距離画像から認識された初期把持位置を元に物体を1個だけ画像外に退避させ、その後生成した第二の距離画像との差となって現れる物体の形状や位置情報を元に現在退避中の物体の正確な把持位置を計算し、ロボットが物体を目的の位置まで移動する際に初期把持位置と正確な把持位置との差を補正するというものである。

【0146】また、この実施の形態9.では、初期把持位置を求める手法として、第一の距離画像から最上段物体領域のコーナ部をもとにコーナ点から一定のオフセットをかけた位置として把持位置を決定したが、例えば距離画像における最上段領域とそれ以外の領域を2値画像とみなして2値テンプレートマッチング法を用いて把持位置を決定してもよい。

【0147】なお、この実施の形態9.では、物体の一時的退避として、ロボットハンドで初期把持位置を把持して画像視野外にハンドで物体を把持したまま移動させる動作を行ったが、一旦所定の場所に物体を下ろしてもよい。下ろした後に再びハンドで物体を把持する際に初期把持位置を把持できれば、この実施の形態9.と同様の動作と考えることができる。

【0148】さらに、この実施の形態9.では退避動作として、物体を把持したロボットハンドを所定の位置で静止させる動作としたが、図25におけるステップST906からST908までの処理を高速に実行することにより、退避動作は継続的にハンドを移動する動作と兼

ねることもできる。即ち、物体の画像視野外に出し、移動作業の目的たる移載場所近辺にロボットハンドを移動している間に位置補正量計算を完了し、移載のためにロボットが物体を所定の場所におく時点で位置補正量を加味した動作を行っても同様の効果を有する。

【0149】

【発明の効果】この発明の請求項1に係る画像処理装置は、複数の物体の距離画像を生成する距離画像生成手段と、距離画像から物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データをもとに物体の2次元画像上での2次元基準パターンを生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものとしたことにより、認識対象物体の画像入力手段、距離画像生成手段が粗い精度のものですむことから装置として小型、安価なものでよく、投光装置もシンプルなものでよく、信頼性が向上する。

【0150】この発明の請求項2に係る物体移載装置は、複数の物体の距離画像を生成する距離画像生成手段と、距離画像から物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データをもとに物体の2次元画像上での2次元基準パターンを生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えた画像処理装置を物体移載装置に組み合わせたものであり、このようにしたことにより、物体移載装置として、認識対象物体の画像入力手段、距離画像生成手段が粗い精度のものですむことから装置として小型、安価なものでよく、投光装置もシンプルなものでよく、信頼性の高い物体移載装置が得られる。

【0151】この発明の請求項3に係る画像処理装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段とで構成したので、認識対象物体に対してランダムドットパターンが投光されるため、表面に図柄のない無地

の物体、あるいは表面に光が反射するラベルが貼ってあっても、表面各位値のにおける距離データを正確に得られ、投光器の構成もシンプルで、小型で安価なものでよく信頼性が高くなる。

【0152】この発明の請求項4に係る物体移載装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段で構成された画像処理装置を物体移載手段に組み合わせたものであり、このようにしたことにより、認識対象物体に対してランダムドットパターンが投光されるため、表面に図柄のない無地の物体、あるいは表面に光が反射するラベルが貼ってあっても、表面各位値が正確に得られ、物体移載装置としての、物体の検出時間が短く、物体の位置は正確に検出され、物体の移載は素早く正確に行うことができ、投光器の構成も、小型、安価なものとなり、信頼性が向上する。

【0153】この発明の請求項5に係る画像処理装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されたものとしたので、少ない画像メモリで、短い演算時間で、高解像度の距離画像が生成され、物体候補の抽出精度も高くすることができる効果がある。

【0154】この発明の請求項6に係る物体移載装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成された画像処理装置を、物体移載手段に組み合わせたものであり、短い演算時間で、高解像度の距離画像が得られ、物体候補の抽出時間が短く、抽出精度が高く、物体の位置は正確に検出され物体の移載は素早く正確に行うことができる。

【0155】この発明の請求項7に係る画像処理装置は、ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックの対に関して高信頼に類似度を計算する高信頼類似度計算手段と、高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成されたもので、距離画像を生成するようにしたので、高速度で高精度の距離画像が得られ、画

像対応づけの信頼性が高くなり、装置全体の信頼性が高くなる。

【0156】この発明の請求項8に係る物体移載装置は、ブロック対応付手段が、左右の対応画像ブロックを高速に探索する高速類似度計算手段と、探索結果であるブロックのペアに関して高信頼に類似度を計算する高信頼類似度計算手段と、高信頼類似度が一定しきい値以下であれば対応ブロック無しと判定する対応ブロック信頼度判定手段とで構成された画像処理装置を、物体移載手段に組み合わせたものであり、高速度で高精度の距離画像が得られるようにしたので、物体候補の抽出時間が短く、抽出精度が高く、画像対応づけの信頼性が高くなり、物体の位置は正確に検出され物体の移載は素早く正確に行うことができる。

【0157】この発明の請求項9に係る画像処理装置は、3つ以上の画像入力手段を有し、ブロック対応付手段が、3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成するようにしたので、ステレオ画像の対の欠落ブロックが生じててもステレオ画像の組合せを換えて補完されるので情報の欠落が防止され精度の高い距離画像が得られる効果がある。

【0158】この発明の請求項10に係る物体移載装置は、3つ以上の画像入力手段を有し、ブロック対応付手段が上記3つ以上の画像入力手段から得られるそれぞれの画像の内2つの画像を選択した画像ペアを2種類以上選択し、ブロック対応付を行ない、それぞれの対応付探索結果を統合して距離画像を生成する画像処理装置を物体移載手段に組み合わせたものとしたので、物体の位置が精度よく検出され、移載は正確に行うことができる。

【0159】この発明の請求項11に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列挙生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものとし、距離画像から抽出された最上段面領域に対して、対象物体の配置仮説から複数の候補物体の位置を検出するので、移載手段への情報伝達は1回でよく、通信時間が短く、予めきめられた積載方法による場合は、まちがって積載していることも検出でき、さらに、最上段物体のどれから把持するかを決定することもでき



る。

【0160】この発明の請求項12に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置と物体移載手段と組み合わせ、距離画像から抽出された最上段面領域に対して、対象物体の配置仮説から複数の候補物体の位置を検出するので、移載手段への情報伝達は1回でよく、通信時間が短く、予めきめられた積載方法による場合は、まちがって積載していることも検出でき、さらに、最上段物体のどれから把持するかを決定することもできる。

【0161】この発明の請求項13に係る画像処理装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有するものとしたので、認識物体の積みパターンが基準となる積みパターンになっていない場合に装置使用者や上位システムの管理者に正しく積まれていることを警告し危険な状態を回避することができる。

【0162】この発明の請求項14に係る物体移載装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と移載手段とを組み合わせたものとしたので、認識物体の積みパターンが基準となる積みパターンになっていない場合に装置使用者や上位システムの管理者に正しく積まれていることを警告し危険な状態を回避することができる。

【0163】この発明の請求項15に係る画像処理装置は、物体配列の基準となるデータを格納する物体寸法データベースから自動的に生成される人工基準パターンを用い、対象となる物体の位置を検出して、検出位置にお

いて前記対象物体に相当する画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては前記実画像基準パターンを利用するように動作するように構成したものであり、画像メモリは小さな容量のものでよく、画像入力時間が短くなり、2回目以降の物体がより早く正確に検出できる。

【0164】この発明の請求項16に係る物体移載装置は、物体配列の基準となるデータを格納する物体寸法データベースから自動的に生成される人工基準パターンを用い、対象となる物体の位置を検出して、検出位置において対象物体に相当する画像領域を切り出す実画像基準パターン切り出し手段と、切り出された実画像基準パターンを格納する実画像基準パターン格納手段とを有し、以降の認識動作においては実画像基準パターンを利用するように動作するように構成した画像処理装置と移載手段を組み合わせたものとしたので、画像入力時間が短く物体が正確に検出・移載することができる物体移載装置となる。

【0165】この発明の請求項17に係る画像処理装置は、段積みされた複数の対象物体の距離画像を生成する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、この第一の距離画像生成手段に格納された距離画像の情報をもとに、物体の把持位置を計算する初期把持位置計算手段と、一時的に物体を所定の位置に退避させた後に、残りの積荷の距離画像を格納する第二の距離画像格納手段と、第一の距離画像格納手段に格納された距離画像と、第二の距離画像格納手段に格納された距離画像とを比較する距離画像比較手段と、この距離画像比較手段の比較結果を基に退避した物体の把持位置を決定し、この把持位置と初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有するものとしたので、距離画像のみで物体の位置認識が可能であり、濃淡やエッジ画像が使いにくい物体に対しても適用でき処理時間も短くなる。

【0166】この発明の請求項18に係る物体移載装置は、段積みされた複数の対象物体の画像入力手段からの距離分布を計測する距離画像生成手段と、生成された距離画像を格納する第一の距離画像格納手段と、物体を把持し移動する物体移動手段と、上記距離画像の情報をもとにして物体の把持位置を計算する初期把持位置計算手段と、物体移載手段によって物体を把持位置で把持し、一時的に所定の位置に対比させた後に残りの積荷の距離画像を格納する第二の距離画像格納手段と、第一の距離画像格納手段に格納された距離画像と、第二の距離画像格納手段に格納された距離画像を比較する距離画像比較手段と、距離画像の比較結果を使用して退避した物体の把持位置を決定し初期把持位置との差を位置補正值として検出する位置補正量検出手段とを有する画像処理装置

と物体を把持し移載する物体移載手段とを組み合わせたものであり、距離画像のみで物体の位置認識ができ、濃淡やエッジ画像が使いにくい物体に対しても正確に認識して安全に適正位置に移載できる物体移載装置となる。

【図面の簡単な説明】

【図 1】 この発明の実施の形態 1. の構成を示すブロック図である。

【図 2】 実施の形態 1. の動作の流れを示すフローチャートである。

【図 3】 実施の形態 1. における各段階において生成された画像である。

【図 4】 この発明の実施の形態 2. の構成を示すブロック図である。

【図 5】 実施の形態 2. の動作の流れを示すフローチャートである。

【図 6】 ランダムドットパターン照射説明図である。

【図 7】 ステレオ画像のブロックマッチングの説明図である。

【図 8】 この発明の実施の形態 3. の構成を示すブロック図である。

【図 9】 実施の形態 3. の動作の流れを示すフローチャートである。

【図 10】 低解像度距離画像から段差領域部を抽出し、高解像度距離画像の段差領域部をのみ高解像度のステレオ対応を行う状況の説明図である。

【図 11】 この発明の実施の形態 4. の構成を示すブロック図である。

【図 12】 実施の形態 4. の動作の流れを示すフローチャートである。

【図 13】 この発明の実施の形態 5. の構成を示すブロック図である。

【図 14】 実施の形態 5. の動作の流れを示すフローチャートである。

【図 15】 ステレオブロック対応付動作のフローチャートである。

【図 16】 ステレオ画像のステレオ対応付の様子を模式的に示す図である。

【図 17】 この発明の実施の形態 6. の構成を示すブロック図である。

【図 18】 実施の形態 6. の動作の流れを示すフローチャートである。

【図 19】 物体候補仮説を用いて物体候補を抽出する動作状況の各段階を示す図である。

【図 20】 この発明の実施の形態 7. の構成を示すブロック図である。

【図 21】 実施の形態 7. の動作の流れを示すフローチャートである。

\* 【図 22】 この発明の実施の形態 8. の構成を示すブロック図である。

【図 23】 実施の形態 8. の動作の流れを示すフローチャートである。

【図 24】 この発明の実施の形態 9. の構成を示すブロック図である。

【図 25】 実施の形態 9. の動作の流れを示すフローチャートである。

【図 26】 初期把持位置を補正する動作の各段階を示す図である。

【図 27】 従来の画像処理装置の処理手順を示す説明図である。

【図 28】 図 27 に示す処理の動作のフローチャートである。

【図 29】 パターン光を用いた空間コード化の原理を示す図である。

【図 30】 他の従来例の距離画像計測システムの説明図である。

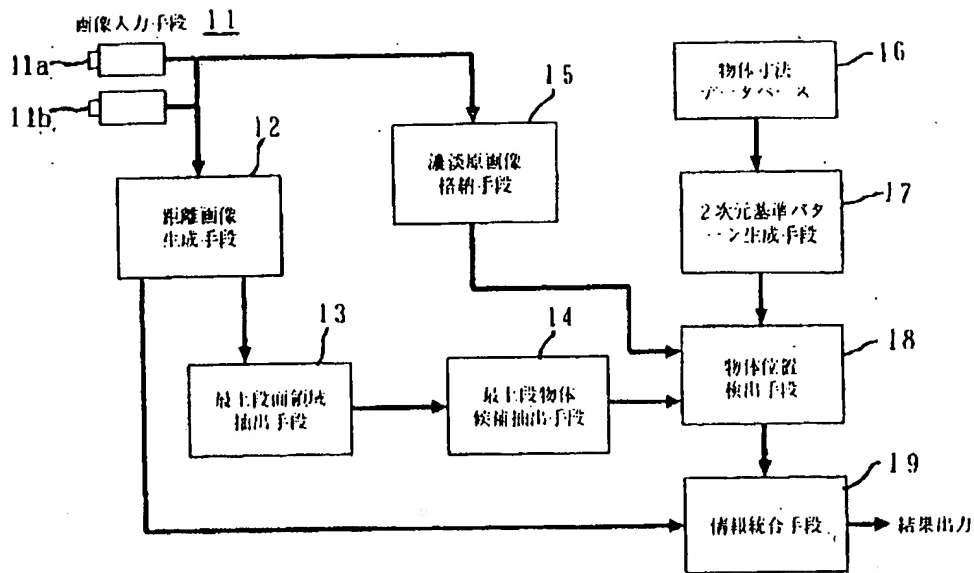
【図 31】 従来の別の画像処理のステレオ対応探索装置の説明図である。

【図 32】 従来の別の画像処理図 30 のフローチャートである。

【符号の説明】

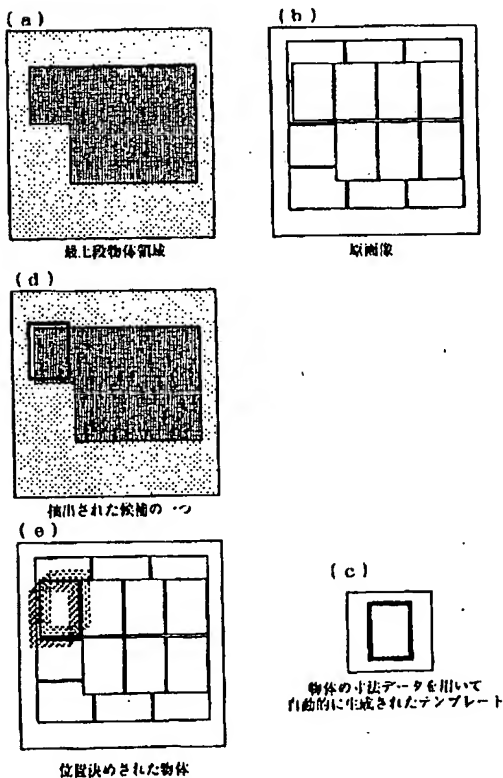
11 画像入力手段、12 距離画像生成手段、13 最上段面領域抽出手段、14 最上段物体候補抽出手段、15 濃淡原画像格納手段、16 物体寸法データベース、17 2次元基準パターン生成手段、18 物体位置検出手段、19 情報統合手段、20 ランダムテクスチャパターン投光手段、21 画像入力手段、22 ステレオ画像ブロック対応付手段、30 パターン投光器、31 低解像度ブロック対応付手段、32 段差領域抽出手段、33 高解像度ブロック対応付手段、34 距離画像合成手段、41a 第一の画像格納手段、41b 第二の画像格納手段、42 高速類似度計算手段、43 高信頼類似度計算手段、44 対応ブロック信頼度判定手段、45 距離画像格納手段、51a 第一の画像入力手段、51b 第二の画像入力手段、51c 第三の画像入力手段、52 ステレオ画像ブロック対応付手段、64 物体候補組合せ仮説生成手段、65 仮説妥当性検証手段、69 物体位置検出手段、71 物体配列基準データベース、72 物体配列判定手段、73 警告発生手段、81 人工基準パターン生成手段、82 実画像基準パターン切り出し手段、83 実画像基準パターン格納手段、88 物体位置検出手段、91 距離画像生成手段、92a 第一の距離画像格納手段、92b 第二の距離画像格納手段、93 距離画像比較手段、94 初期把持位置計算手段、95 位置補正量計算手段、96 物体移動手段。

【図1】

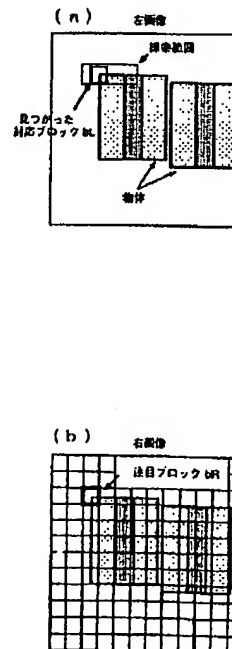


- 11 : 画像入力手段  
 12 : 距離画像生成手段  
 13 : 最上段面領域抽出手段  
 14 : 最上段物体候補抽出手段  
 15 : 濃淡原画像格納手段  
 16 : 物体寸法データベース  
 17 : 2次元基準パターン生成手段  
 18 : 物体位置検出手段  
 19 : 情報統合手段

【図3】

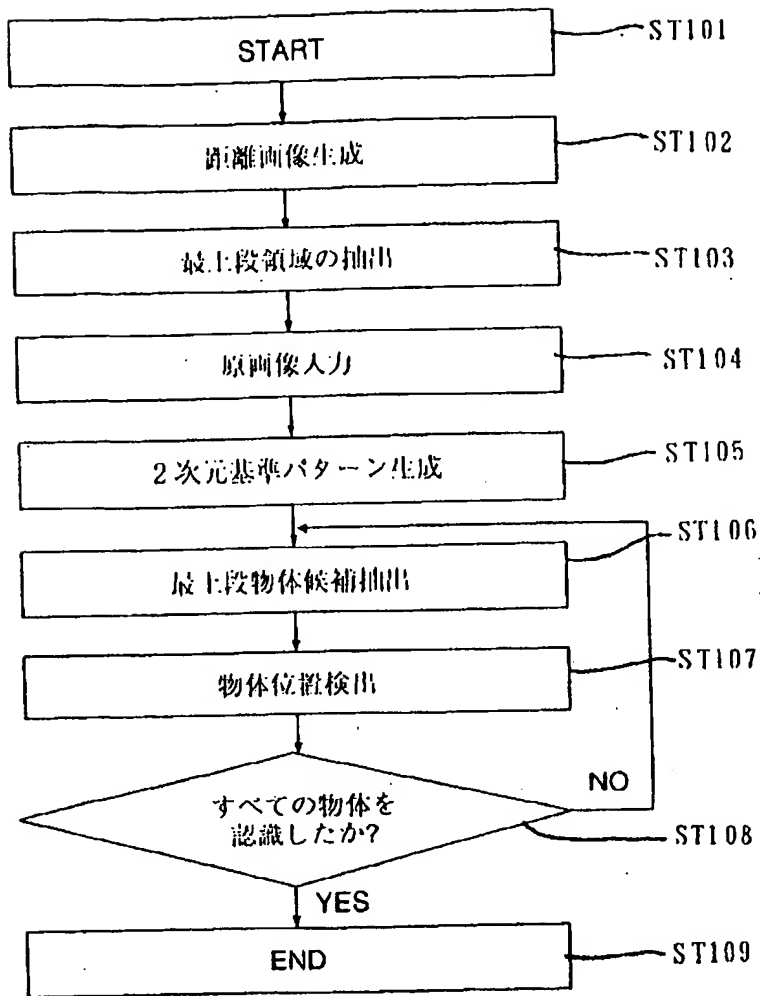


【図7】

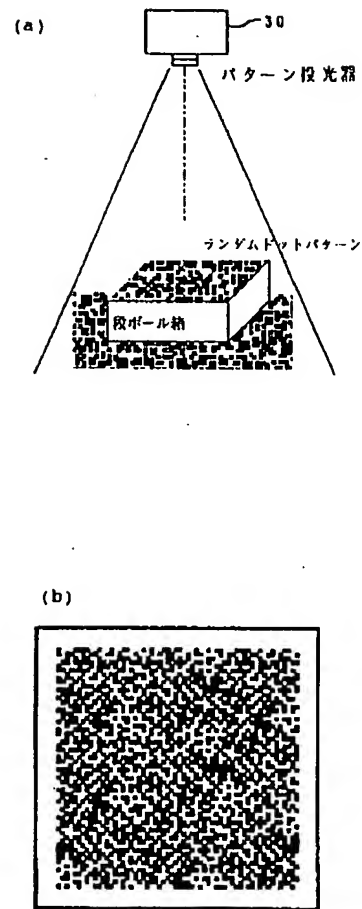




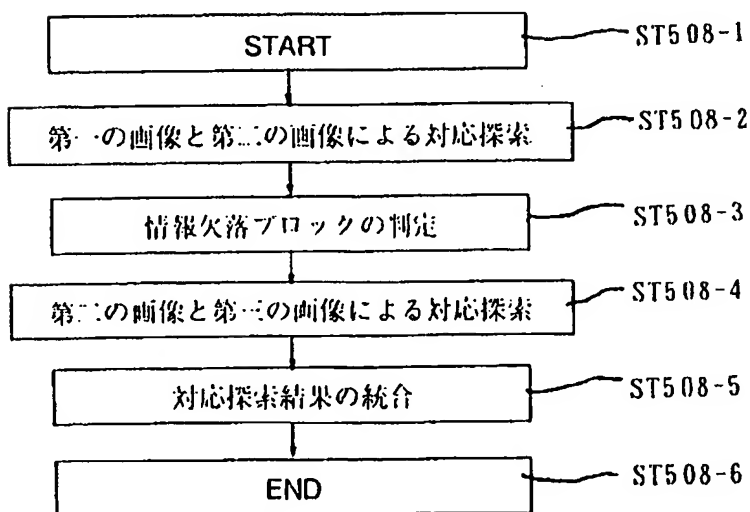
【図2】



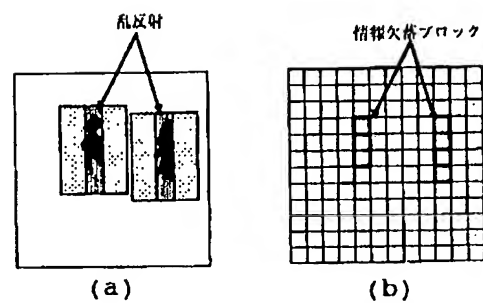
【図6】



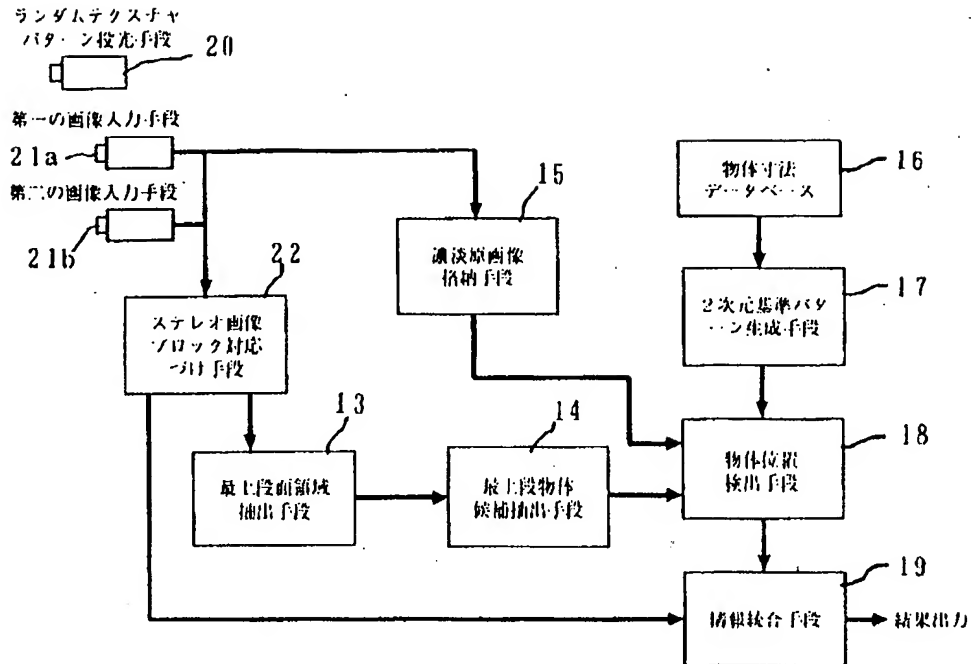
【図15】



【図16】

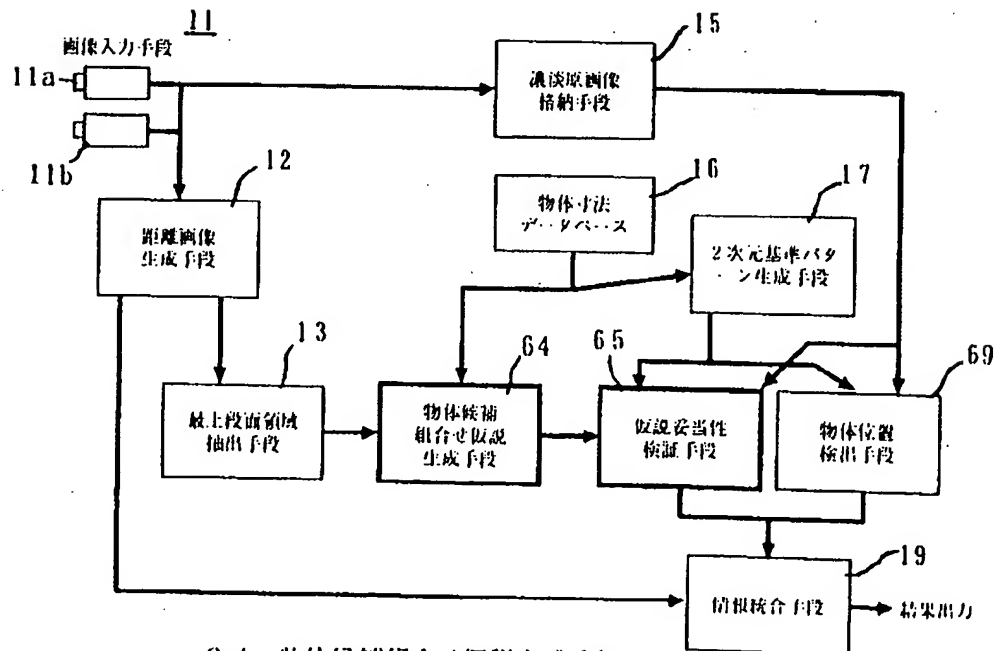


【図4】



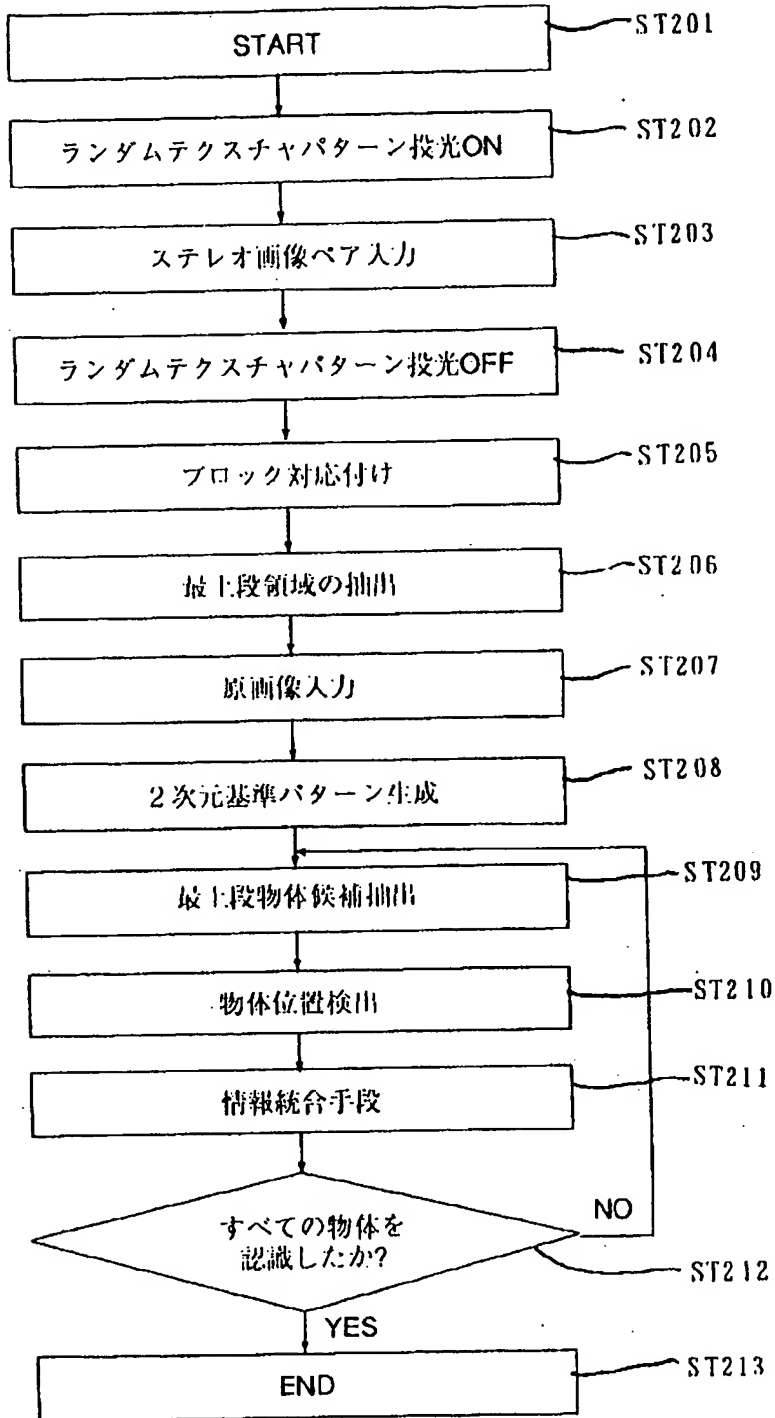
20 : ランダムテクスチャパターン投光手段  
 21 : 画像入力手段  
 22 : ステレオ画像ブロック対応付手段

【図17】

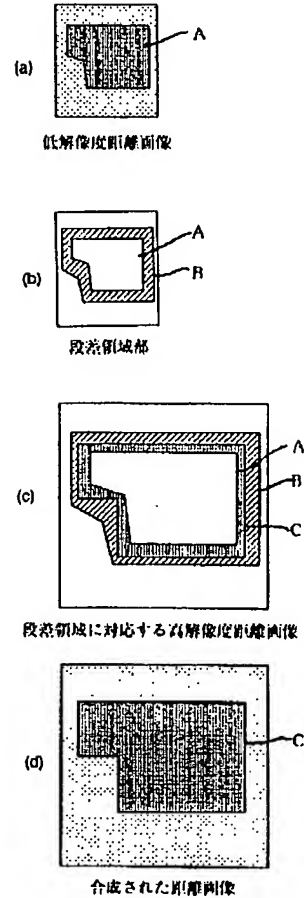


64 : 物体候補組合せ仮説生成手段  
 65 : 仮説妥当性検証手段  
 69 : 物体位置検出手段

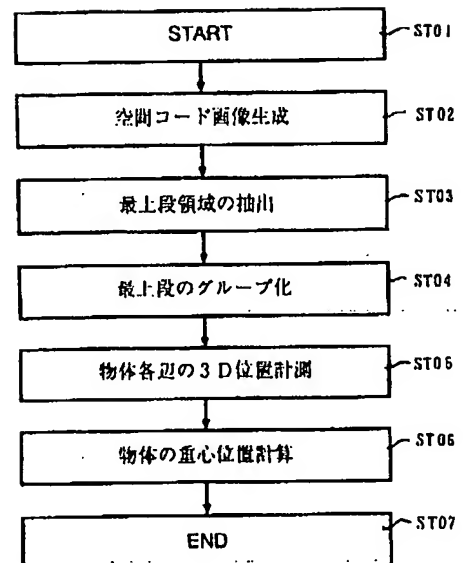
【図5】



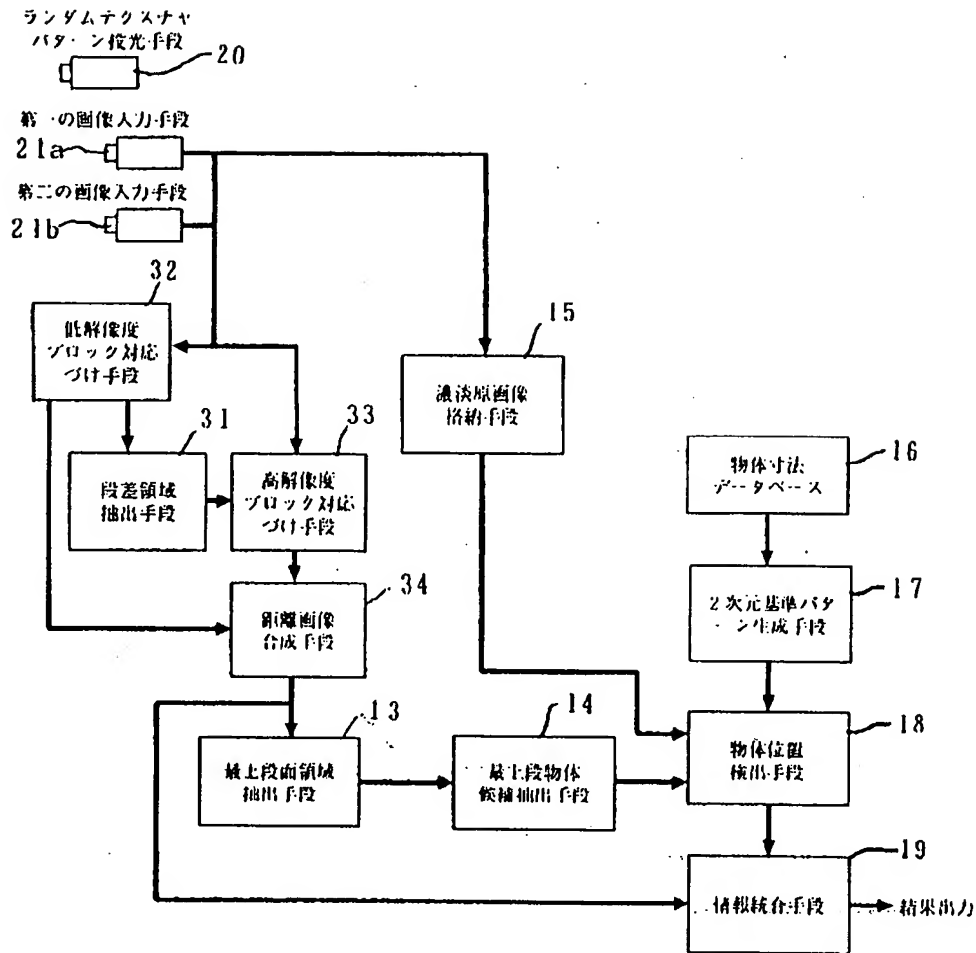
【図10】



【図28】

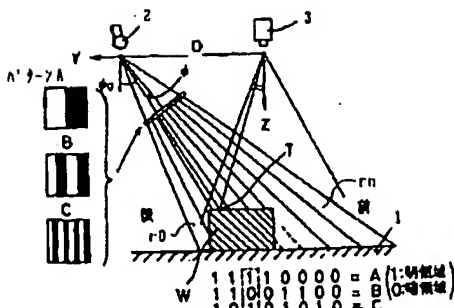


【図8】

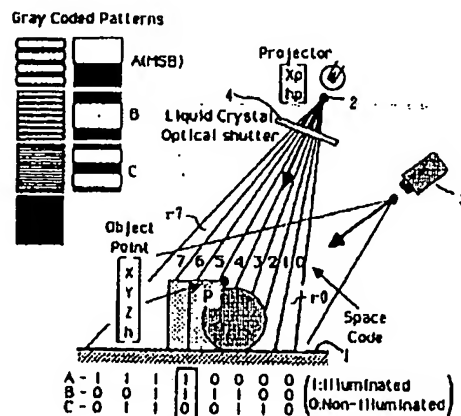


- 31 : 低解像度ブロック対応付手段  
 32 : 段差領域抽出手段  
 33 : 高解像度ブロック対応付手段  
 34 : 距離画像合成手段

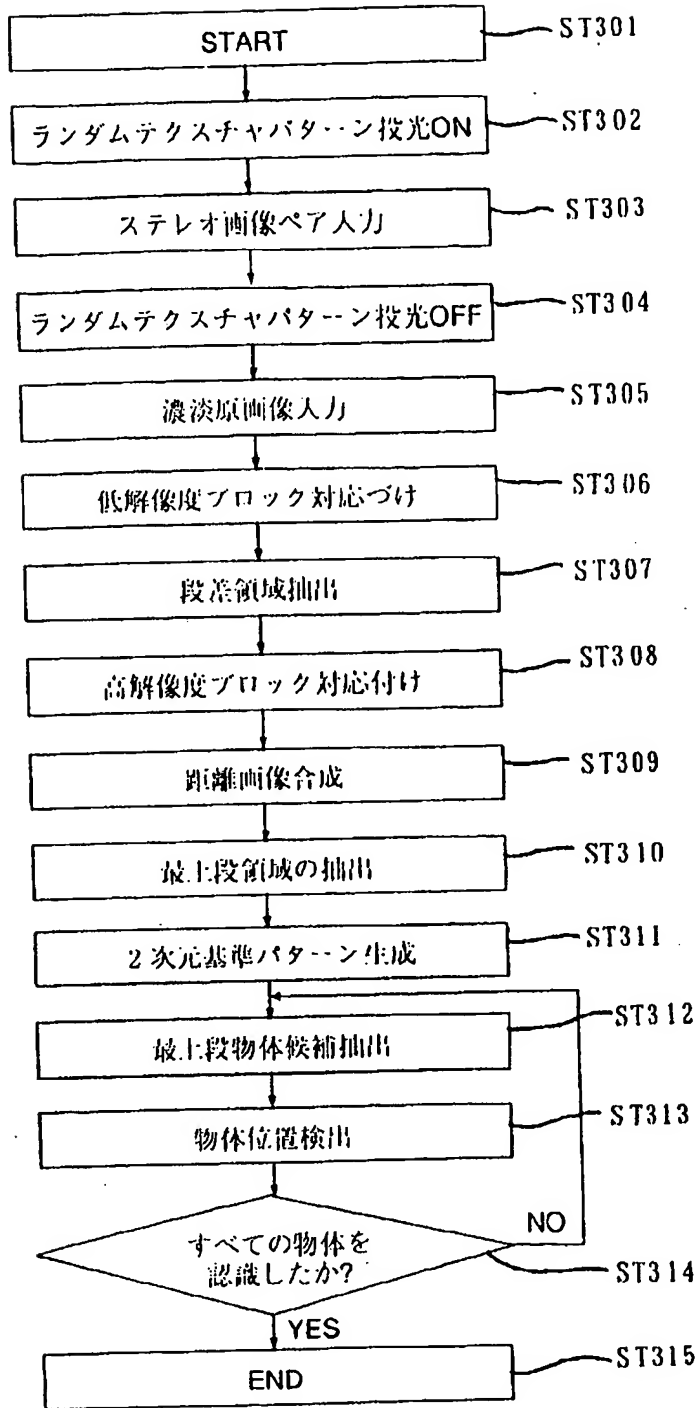
【図29】



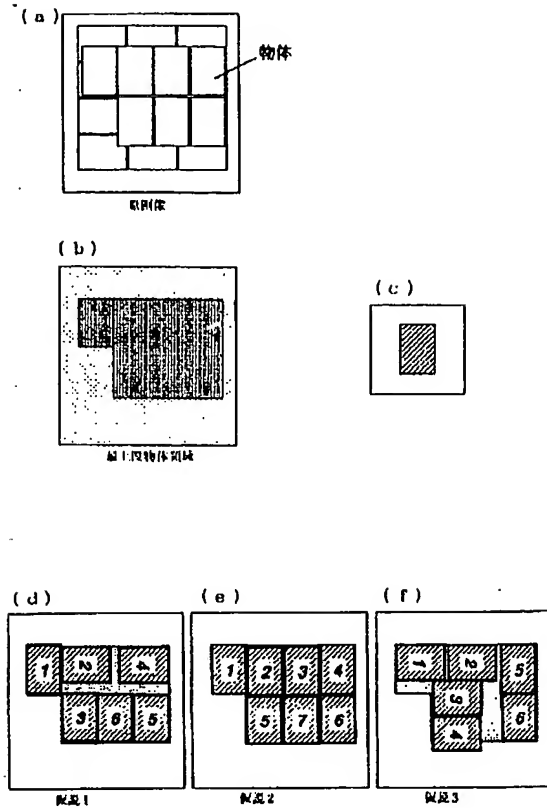
【図30】



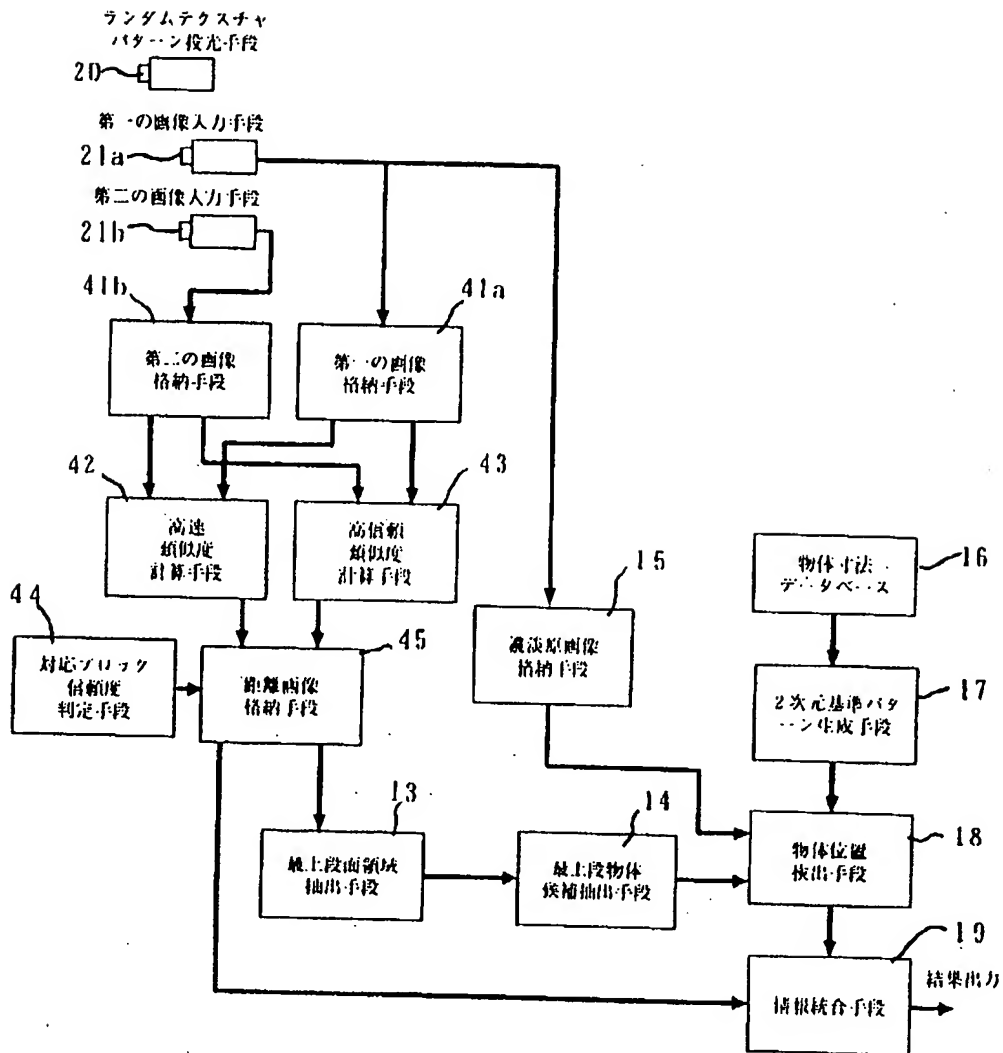
【図9】



【図19】

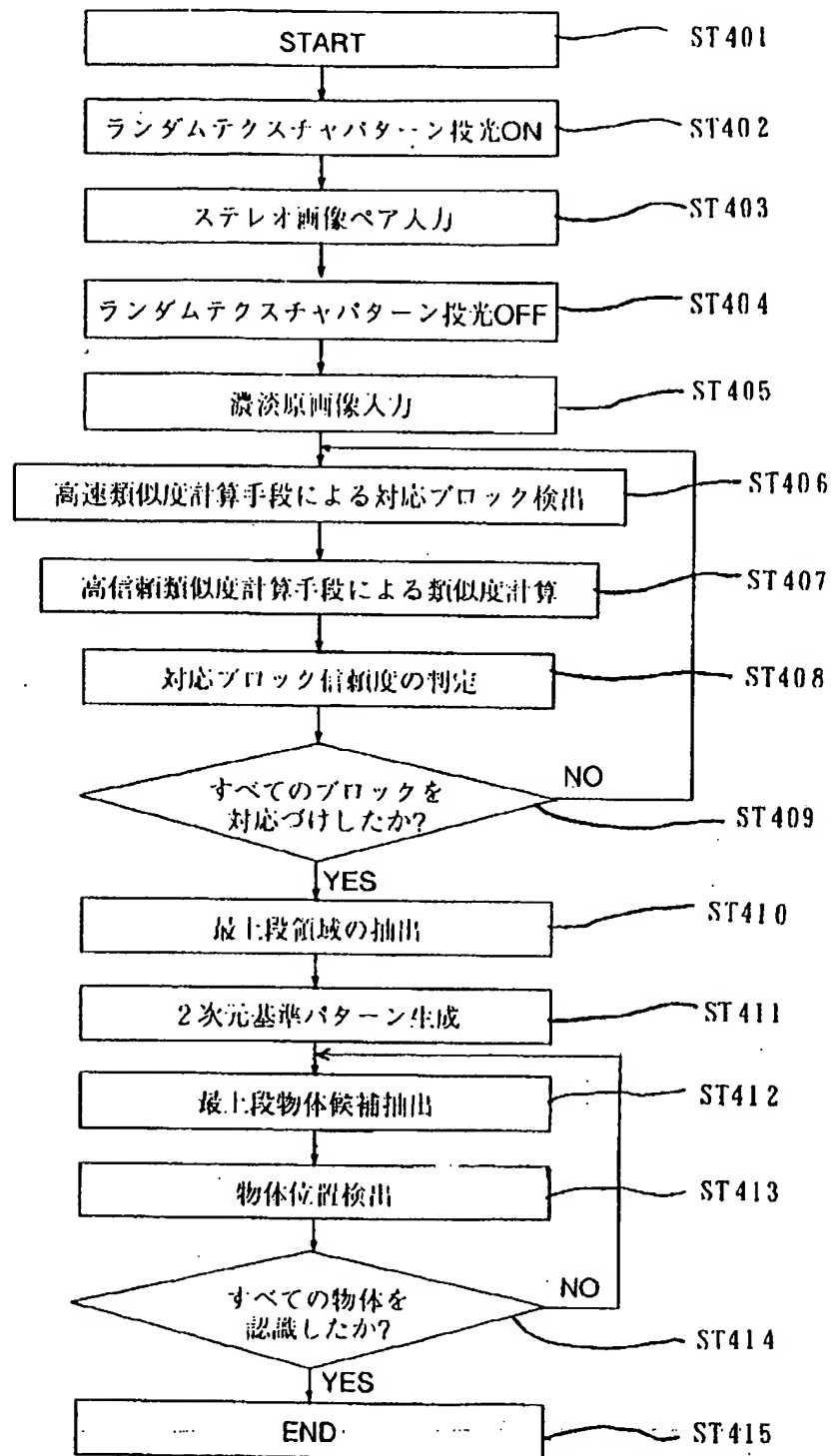


【図1.1】

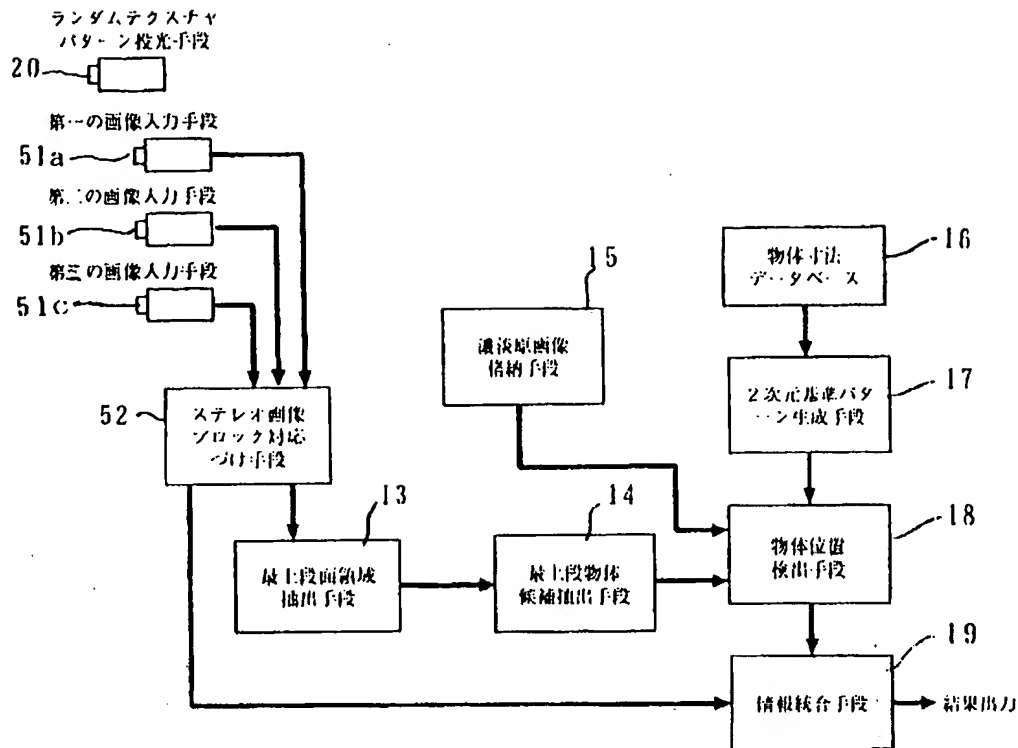


- 41a : 第一の画像格納手段  
 41b : 第二の画像格納手段  
 42 : 高速類似度計算手段  
 43 : 高信頼類似度計算手段  
 44 : 対応ブロック信頼度判定手段  
 45 : 距離画像格納手段

【図12】

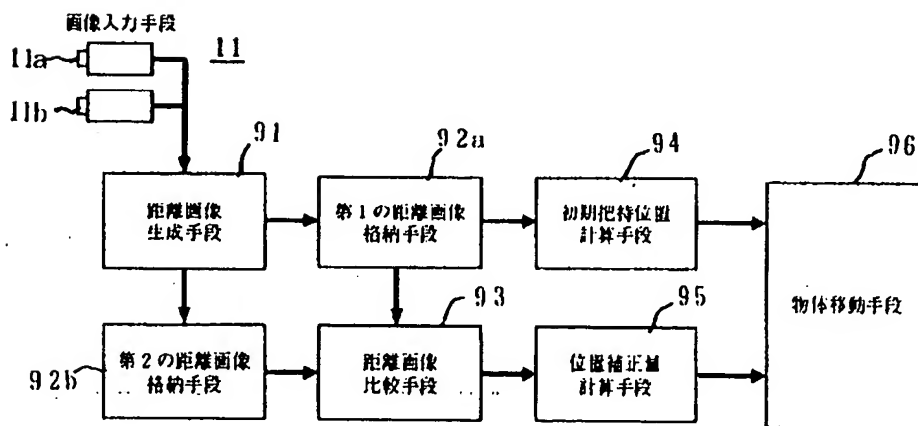


【図13】



51a : 第一の画像入力手段  
 51b : 第二の画像入力手段  
 51c : 第三の画像入力手段  
 52 : ステレオ画像ブロック対応付手段

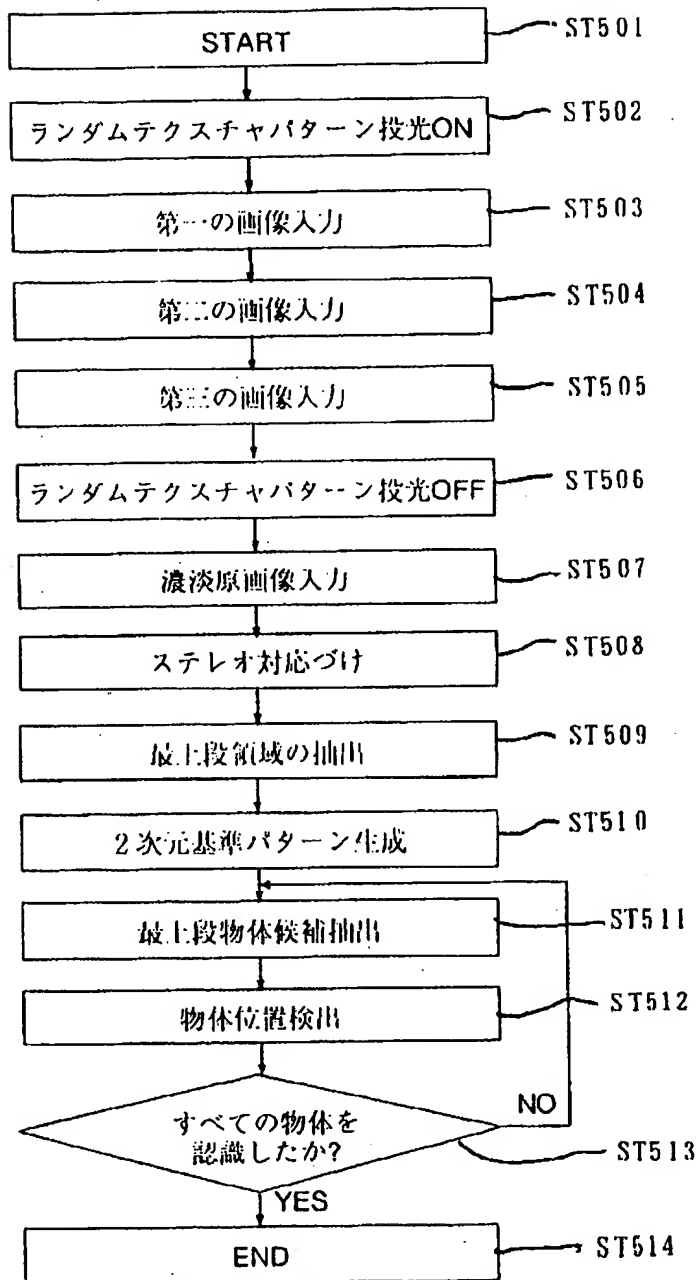
【図24】



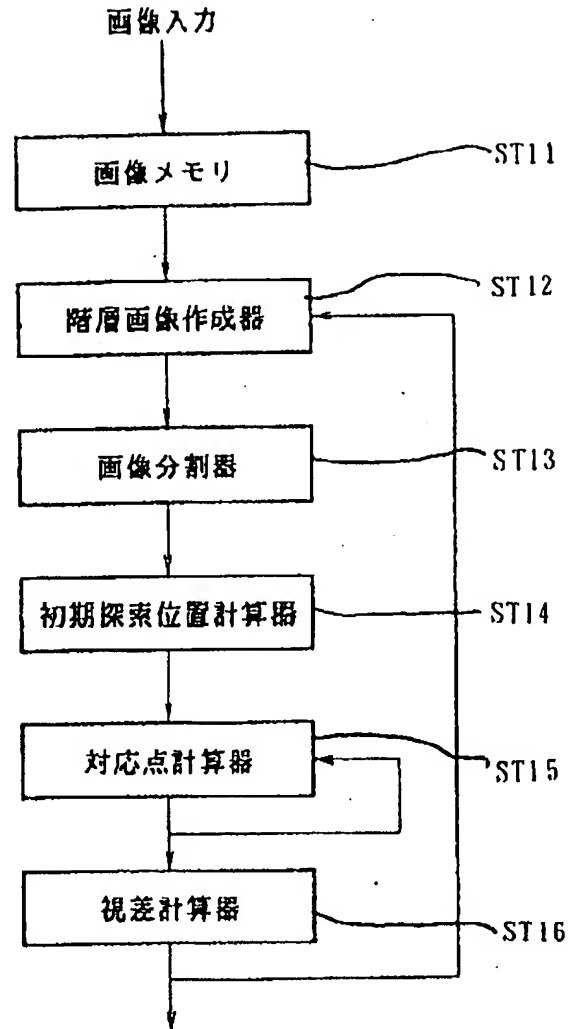
91 : 距離画像生成手段  
 92a : 第一の距離画像格納手段  
 92b : 第二の距離画像格納手段  
 93 : 距離画像比較手段  
 94 : 初期把持位置計算手段  
 95 : 位置補正量計算手段  
 96 : 物体移動手段



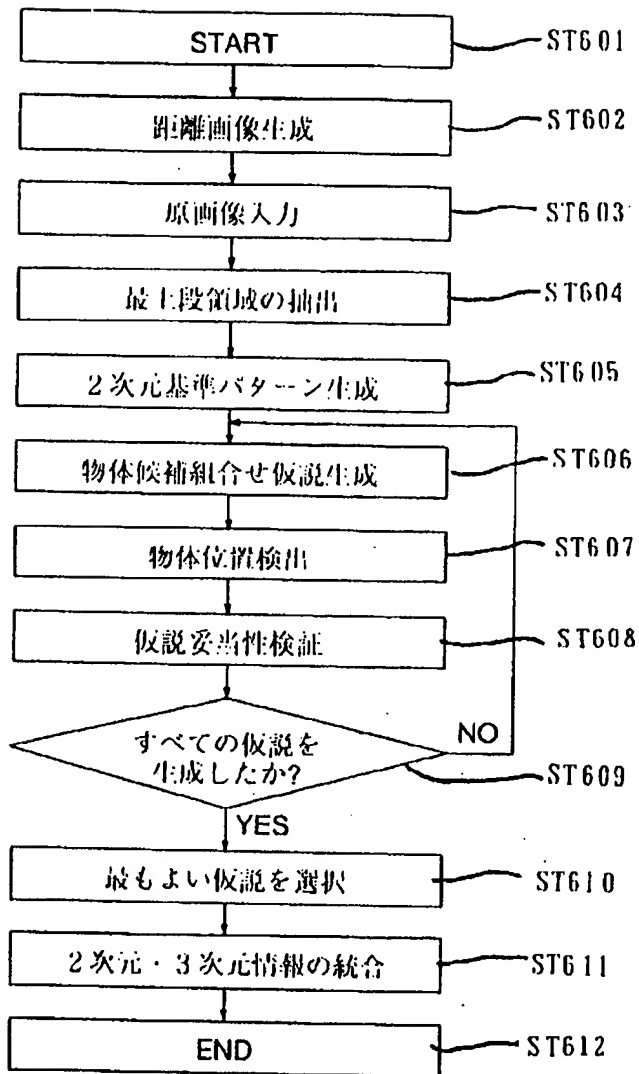
【図14】



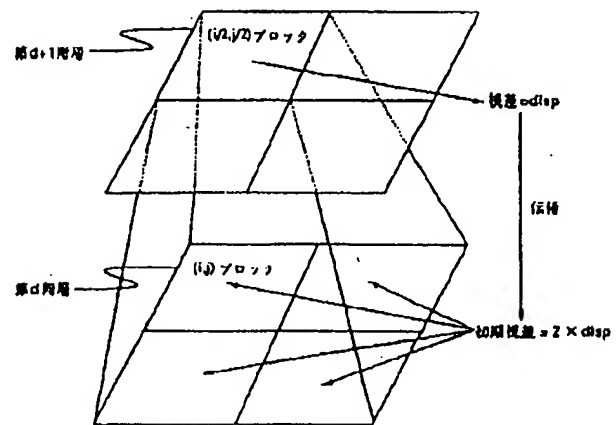
【図32】



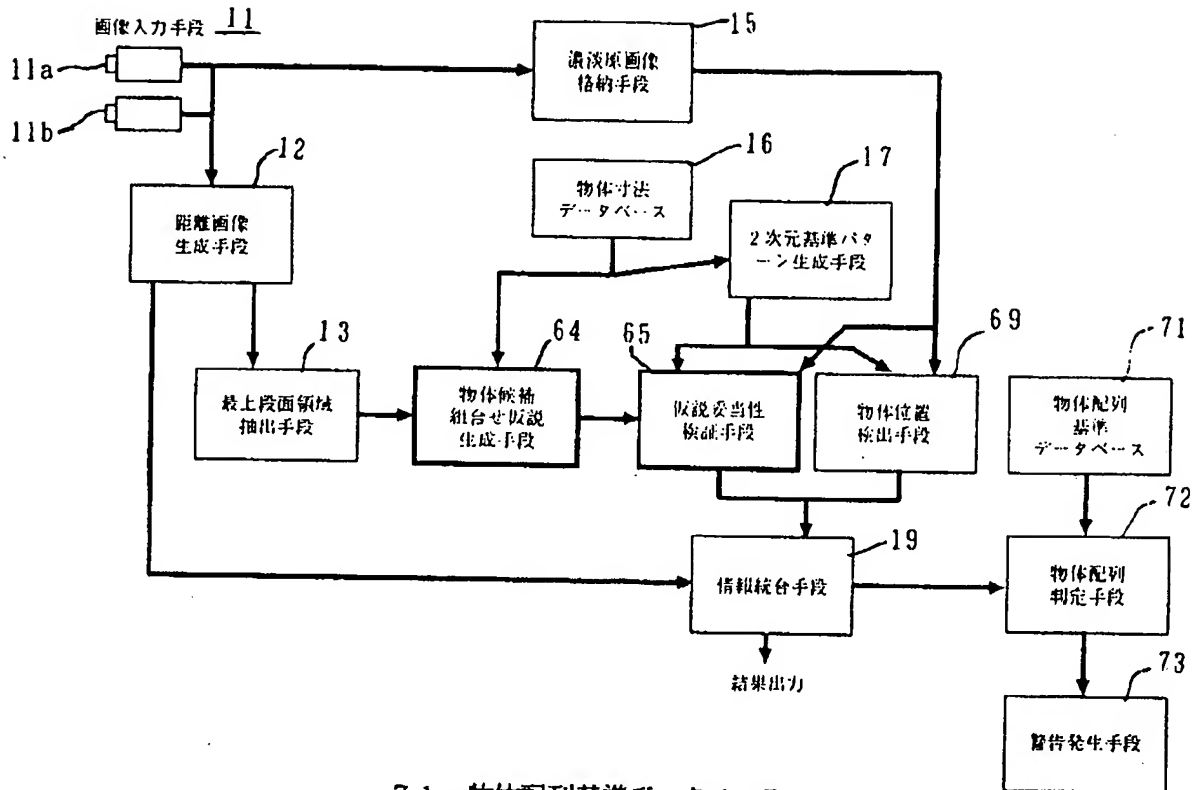
【図18】



【図31】



【図20】

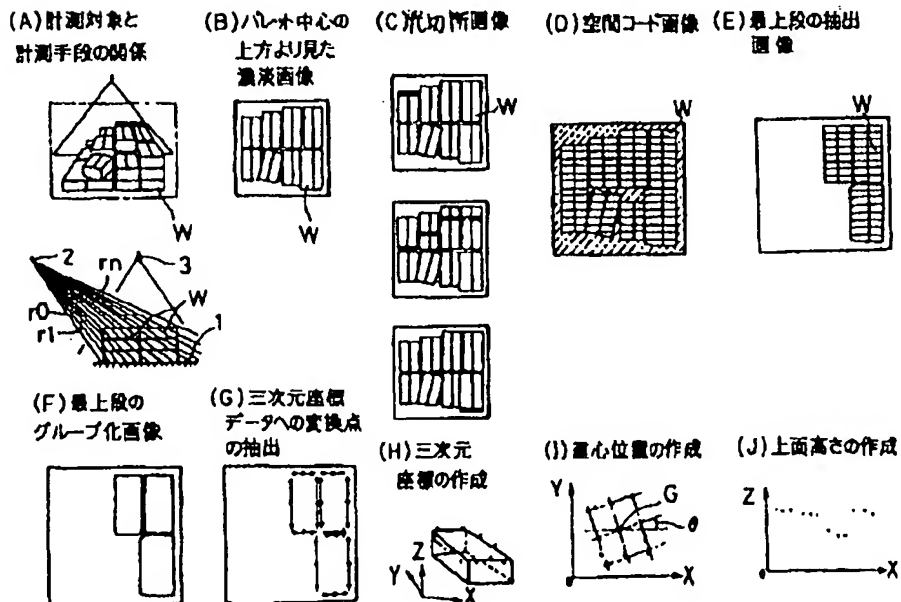


71 : 物体配列基準データベース

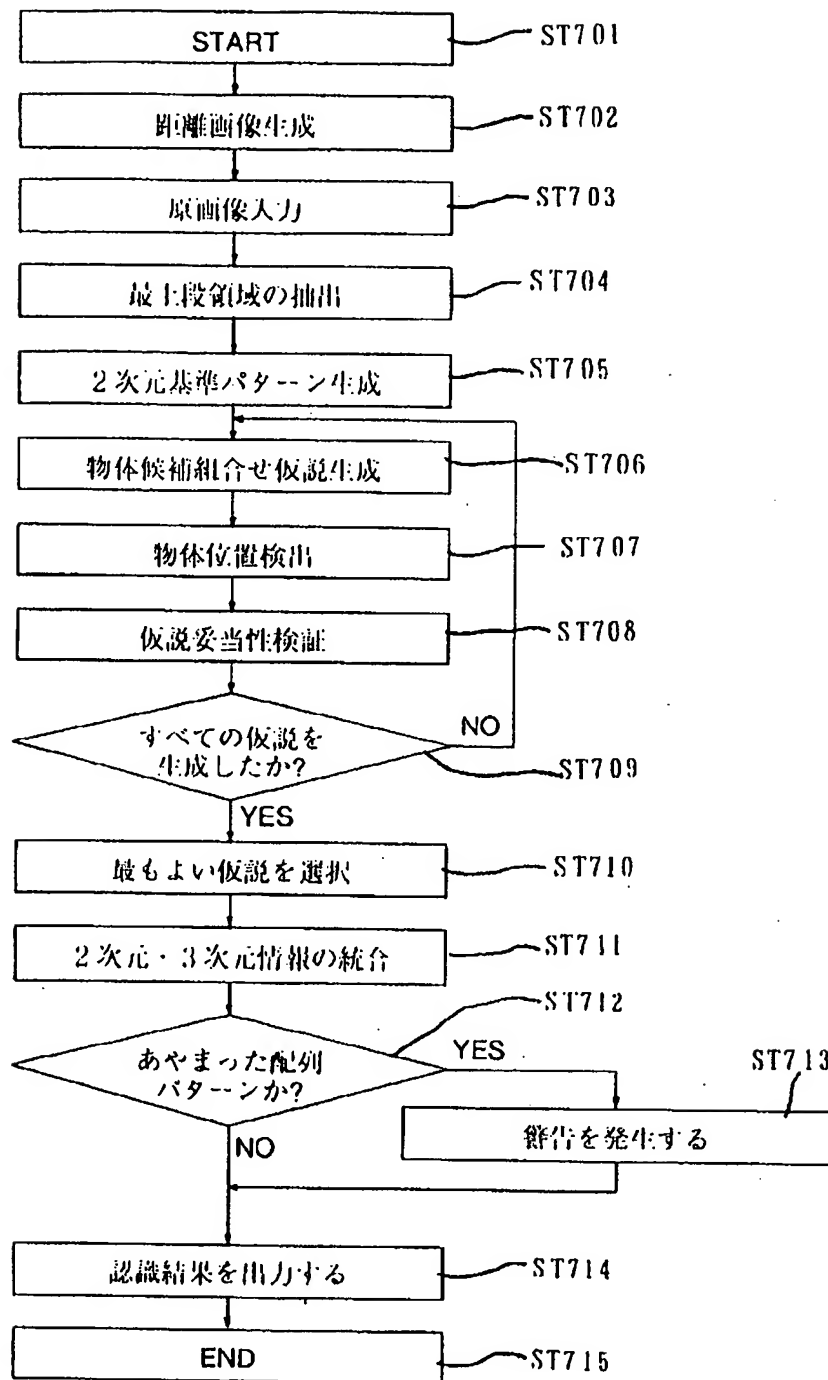
72 : 物体配列判定手段

73 : 警告発生手段

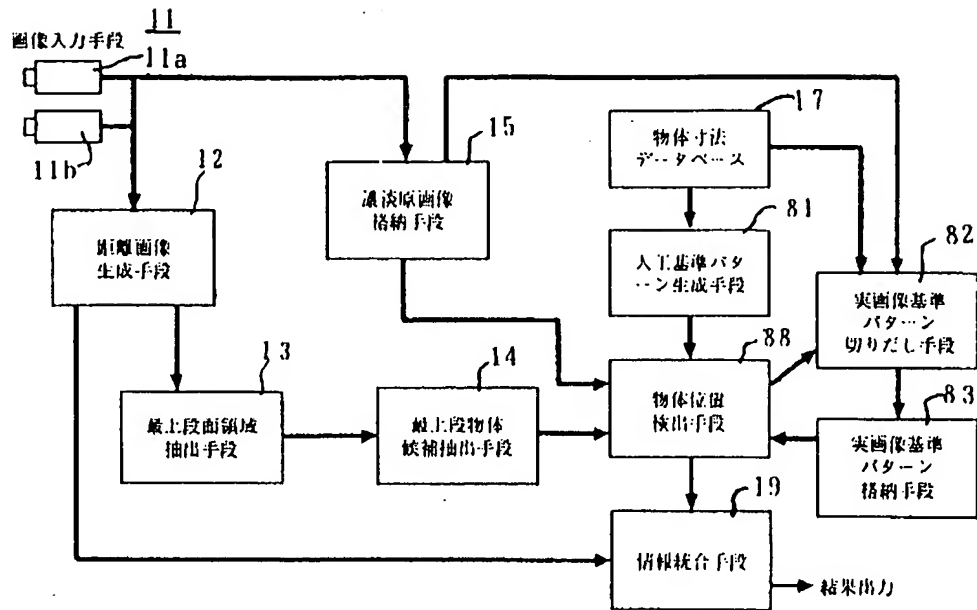
【図27】



【図21】

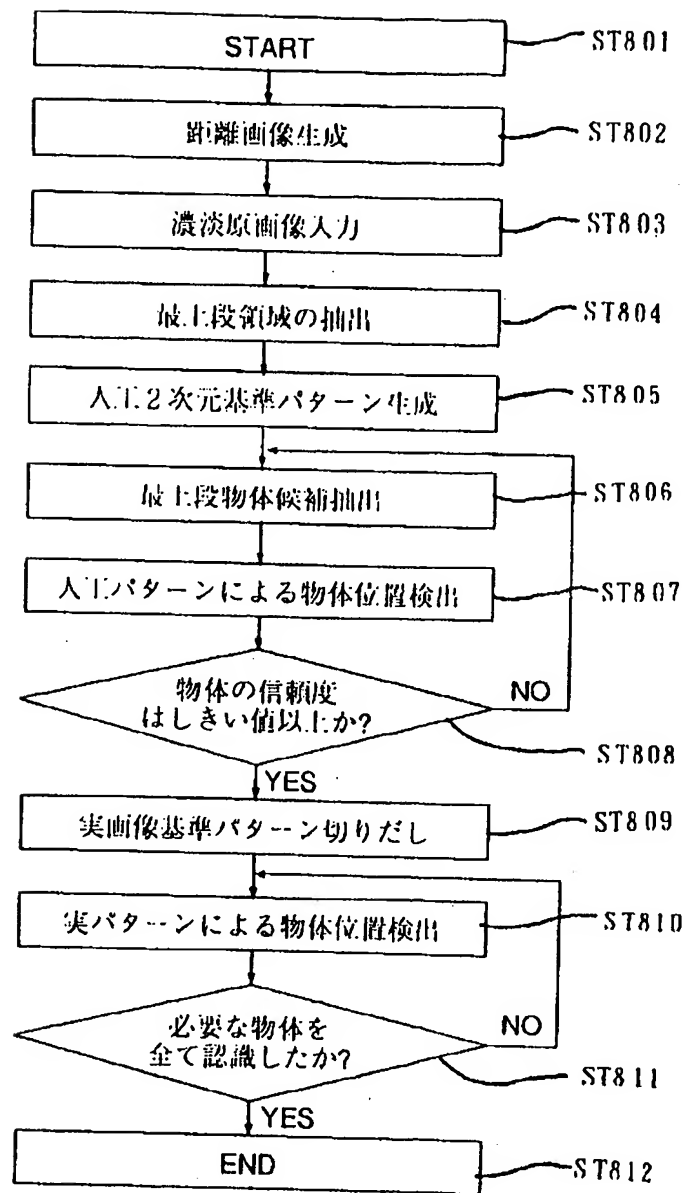


【図22】

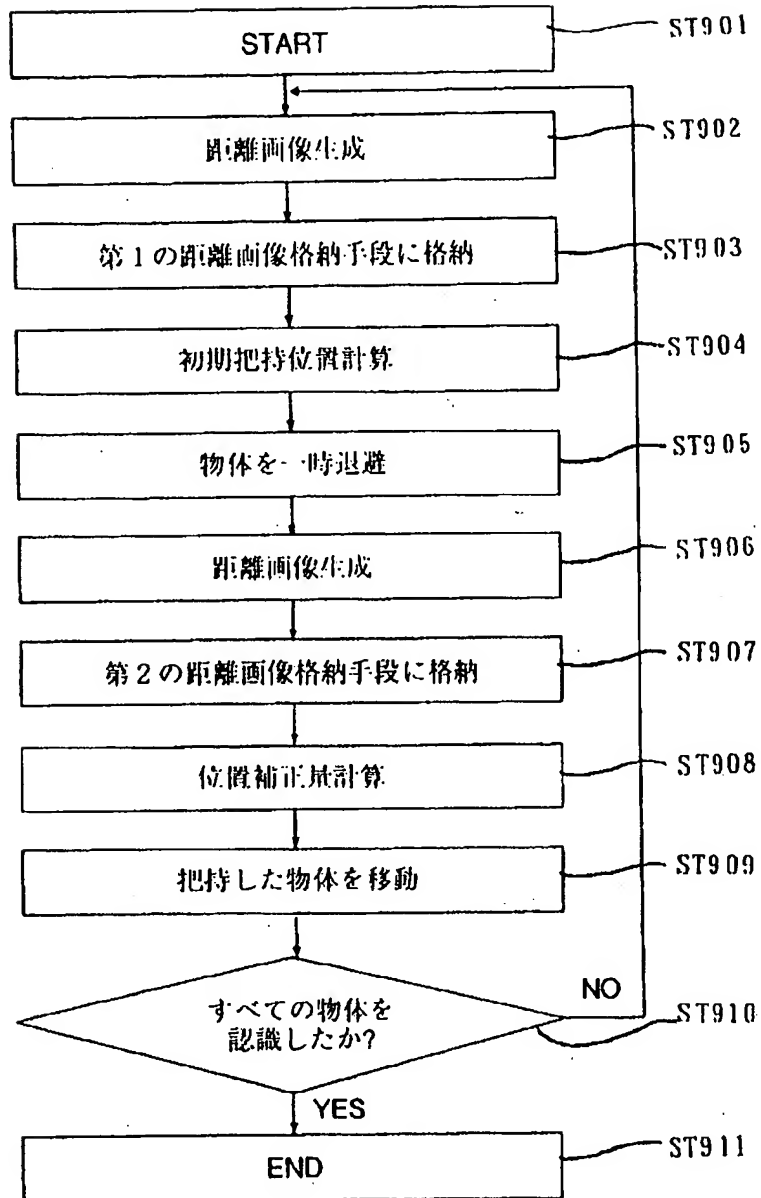


- 81 : 人工基準パターン生成手段  
 82 : 実画像基準パターン切り出し手段  
 83 : 実画像基準パターン格納手段  
 88 : 物体位置検出手段

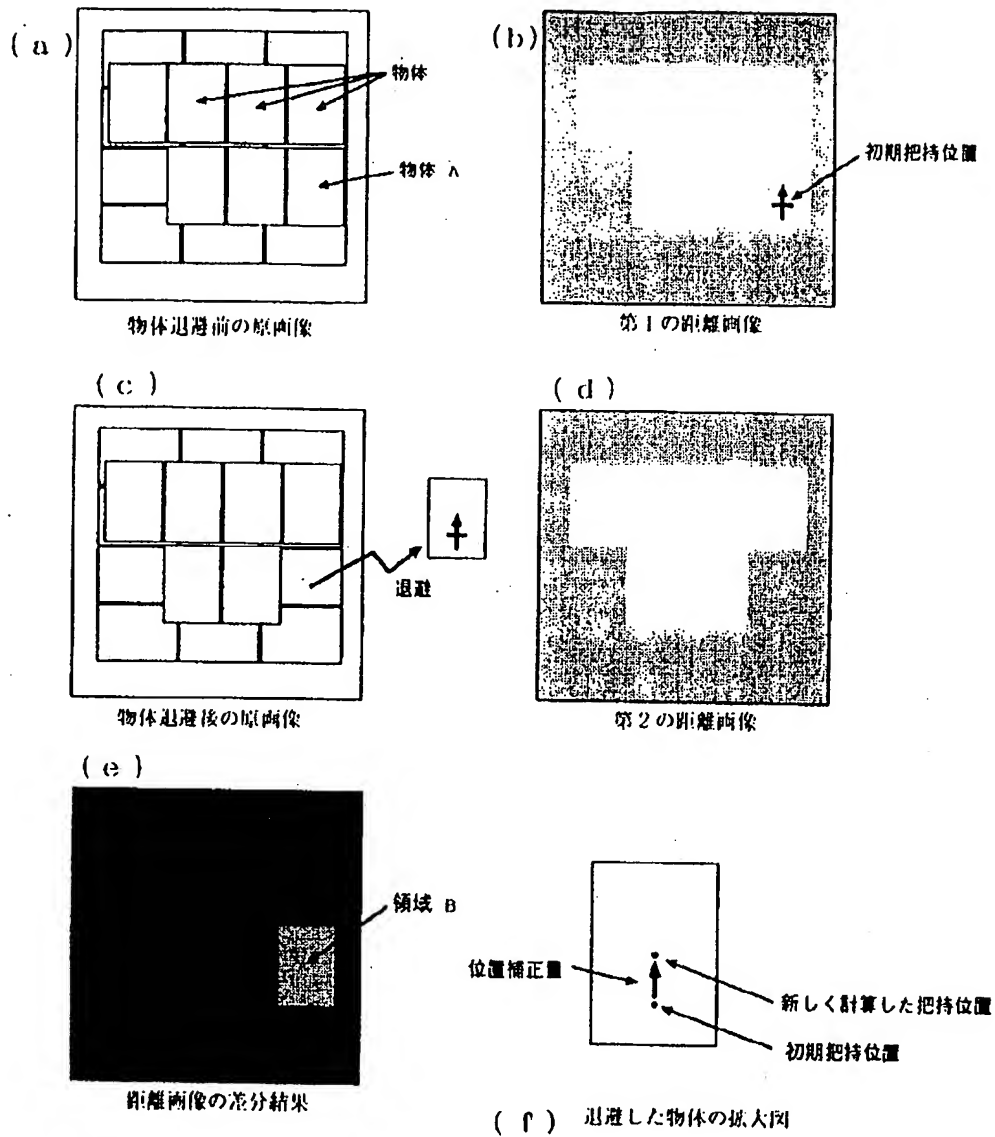
【図23】



【図25】



【図26】





【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成11年(1999)7月30日

【公開番号】特開平10-31742

【公開日】平成10年(1998)2月3日

【年通号数】公開特許公報10-318

【出願番号】特願平8-184917

【国際特許分類第6版】

G06T 7/00

G01B 11/00

G06T 1/00

7/60

【F I】

G06F 15/62 415

G01B 11/00 H

G06F 15/64 M

15/70 350 B

【手続補正書】

【提出日】平成11年1月19日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】画像処理装置および物体移載装置

【特許請求の範囲】

【請求項1】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、上記個別に抽出された最上段物体候補について、上記2次元基準パターンと上記濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、上記各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置。

【請求項2】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを

格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、上記個別に抽出された最上段物体候補について、上記2次元基準パターンと上記濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、上記各手段で得られた物体位置の検出結果と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置および物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する物体移載装置。

【請求項3】 距離画像生成手段は、ランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、上記第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段とで構成されていることを特徴とする請求項1記載の画像処理装置。

【請求項4】 距離画像生成手段が、ランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、上記第一および第二の画像入力手段が撮像した二つの画像間の対応付を行うステレオ画像ブロック対応付手段で構成された画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2記載の物体移載装置。

【請求項5】 距離画像生成手段は、解像度の粗い距離

画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、上記低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されていることを特徴とする請求項1記載の画像処理装置。

【請求項6】 距離画像生成手段は、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、上記低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段で構成された画像処理装置と物体を把持し移載する物体移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項2記載の物体移載装置。

【請求項7】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置。

【請求項8】 積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列举生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置と物体を把持し移載する移載手段を備え、上記画像処理装置の出力信号に基づいて物体を移載する物体移載装置。

【請求項9】 物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証

手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有することを特徴とする請求項7記載の画像処理装置。

【請求項10】 物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と物体を把持し移載する移載手段とを備え、上記画像処理装置の出力信号に基づいて物体を移載する請求項8記載の物体移載装置。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】この発明は、積荷を自動的に移載するロボットや産業機械の積載された荷物の位置を認識する画像処理装置および物体移載装置に関するものである。

##### 【0002】

【従来の技術】積荷を自動的に移載する手段の画像処理装置の例として特開平6-249631号公報に開示されたものがある。図16はその発明の画像計測の処理手順の概略説明図、図17に動作を示すフローチャート、図18はパターン光を用いた空間コード化の原理を示す図である。

【0003】この画像処理装置は、パレット1上に積載された段ボール箱Wを計測対象とするものであり、図18に示すように、パレット1の斜め上方に投光器2、パレット1の中心部上方に配置されたカメラ3からなり、投光器2は各種のパターン光を時系列的に投光することにより、測定空間を互いに積層された楔状測定領域 $r_0 \sim r_n$ に分割するスリットパターン光を投光するものである。スリットパターンは、例えば図18のA、B、Cの白黒の2値パターンの3つのパターン光を投光するようになっており、白は光の当たっている部分、黒は光の当たっていない部分である。このようなパターンは、例えば液晶シャッターのようなドットマトリクス電気シャッターを用いて作ることができる。

【0004】光が当たっている状態を“1”、当たっていない状態を“0”と表現すると、図16においては、パターンAが投光されている時は前から後ろに向かって、半分づつ1、0となる。パターンBでは、前から順に1/4づつ1、0、1、0となる。パターンCでは、1/8づつ1、0、1、0、1、0、1、0となる。

【0005】これら3つのパターンを投光してそれぞれ画像を撮像すると、各画素については「000」から

「111」の3ビットコードにコード化されたデータの補2-

どれかに対応させることができる。このようなコードを空間コードと呼ぶ。ある空間コードに対応するのは3次元空間上では楔状領域であり。そこに物体Wが存在していると物体の表面上の画像領域にコードが割り当てられることになる。一方、カメラで撮像した画像から、画像上のある画素の空間コードがわかれば、どの楔状領域に相当するかがわかるので、三角測量の原理によって物体表面の高さ、即ちカメラからの距離がわかることになる。この説明ではパターンをA、B、Cの3種類としたため、画像全体をつの楔状領域に分割したが、パターンの種類を8種類とすれば、楔状領域の数は256になり、より精密に距離情報の精度を向上させることができる。即ち、空間コード画像を生成することによって、画像上のそれぞれの画素のカメラからの距離が計測できる。

【0006】図16について、図17に示すフローチャートにしたがって説明する。ステップST01でシステムが起動されると、ステップST02において空間コード画像が生成され、図16(A)の状態にある積荷の上面に同図(D)のような空間コード画像が得られ、画像を水平に左から右に走査し、もっとも空間コードの大きいデータを持つ領域を抽出すれば同図(E)のように最上段物体の上面部分荷対応する画像が得られる。ステップST04では、最上段物体領域をグループ化している。空間コード画像において隣接する画素間で空間コードが所定の範囲内に近接している時、それら領域を同じグループと判断したり、各積荷上縁間に生じる溝状の部分やすき間部分の空間コードの変化をもとにグループ化している。グループ化された各物体は、続くステップST05で各物体の3次元位置が計測される。グループ化された物体の辺縁部分に相当する部分の3次元座標を何点かサンプリングし、ステップST06で、それらをもとに物体の重心位置と姿勢、および高さ情報を得る。ステップST07で処理を終了する。このようにして得られた積荷の位置データを用いて、ロボット等で積荷をひとつずつ移載することができる。

【0007】この画像処理装置では、積荷の認識に必要な情報をすべて空間コード画像から獲得し、例えば解像度 $256 \times 256$ または $512 \times 512$ 画素といった高い解像度の距離画像を生成する必要がある、時系列的に投光パターンを変化させる機構を有する高精度のパターン投光器が使用され、時系列で発生させた複数の投光パターンそれぞれに対応させて、画像を入力するので、大容量の画像メモリを必要とする。

【0008】他の従来例として、電子情報通信学会論文誌Vol. J71-D, No. 7 p1240~1257に掲載された佐藤、井口の論文「液晶レンジファインダー—液晶シャッターによる高速距離画像計測システム—」に記載された内容を紹介する。この技術は、上記の図16~図18に示された技術で使用された空間コード

化法による距離画像獲得システムに関するものであり、図19にその原理を示す。液晶シャッター4を用いて時系列的に交番2進コードによるスリットパターンを物体に投影し、カメラ3でそれぞれのパターンに対応する画像を撮像して空間コード画像を生成するものである。パターンとしては図16のような7枚のパターンを用いる。計測の原理は上記図16~図18の従来技術で用いられているものと同一である。

【0009】また別の従来技術として、特開平7-103734号公報に記載された「ステレオ対応探索装置」について説明する。図20は、この発明の階層的ブロックマッチング（疎密探索法）の説明図であり、図21その発明の動作のフローチャートである。

【0010】ステレオ視覚とは、複数の異なる位置に配置された画像入力手段から入力された2つの画像の対象物上の同一点に対応する視差を計測することにより、三角測量の原理からカメラと対象点との距離を計算する手法である。図21のフローチャートのST11において画像が入力され、ST12で入力された2つの画像がそれぞれ画像メモリに格納され、ST13において、格納された画像を階層的により低い解像度に変換する処理が行われて階層画像が生成される。図20は第D階層の画像と第D+1階層の画像との対応関係を模式的に表したものである。この例では、第D階層の画像の1/4領域を第D+1階層の画像としている。

【0011】ステップST14において、画像が小ブロックに分割され、ステップST15において、各ブロックは初期探索位置計算器によって初期探索位置が計算され、ステップST16でパターンマッチングによって他画像におけるブロックの対応点が計算され、ステップST17において、2つの画像間の対応点の位置情報を用い、視差計算器によって視差が計算される。このようにして得られた視差画像をもとに、一段階高解像度の画像を用いて再び視差が計算される。この処理を繰り返して最終的には入力画像と同じ最高解像度の画像に対する対応点探索を行なう。

【0012】

【発明が解決しようとする課題】上記図16~図18に示された画像処理装置では、積荷の認識に必要な情報をすべて空間コード画像から獲得しており、例えば解像度 $256 \times 256$ または $512 \times 512$ 画素といった高い解像度の距離画像を生成する必要がある、高精度のパターン投光器が必要であり、しかもそれは時系列的に投光パターンを変化させる機構を有するものが必要であり、大規模な高価な装置となり、さらに、時系列で発生させた複数の投光パターンそれぞれに対応させて画像を入力する必要がある、大容量の画像メモリも必要とし、大規模な高価な装置となる問題点があった。またそれに要する画像入力時間の総計も多くかかるという問題点もあった。

【0013】上記図19に示す技術においても、空間コード画像のみをもとに積荷を認識するので、図16～図18の場合と同様に、大きな画像メモリも必要とし、大規模な高価な装置となる問題点があった。

【0014】さらに、図16～図18、および図19の画像処理装置は、積荷に必要な情報を全て空間コード画像から獲得するために、積荷において把持したい物体と隣接物体が密に接触している場合、すなわち、稜線の面とり量の小さい段ボール箱においては隣接する物体との境界が三次元的に明確に段差として現れないもの、あるいは、軟らかい内容物が収納されたセメント袋のような袋物の場合は、隣接する物体と密接するため、境界が三次元的には不明確であり、個々の物体を個別に分離して移送手段に把持させることが困難であるという問題点があった。

【0015】図20、図21に示された画像処理装置では、高速にステレオ対応点を探索し、距離画像を短い時間で獲得することを目的としているが、画像中の対象物の如何によらず常に画像上に画一的に設定された小ブロック単位でより粗な解像度の画像から順により密な解像度の画像へと探索を行なうので、あらかじめ高解像度情報が必要な部位がわかっている場合は、逆に処理時間が長くなるという問題点があった。

【0016】この発明は、上記問題点を解決するたもになされたものであり、大容量の画像メモリを必要とせず、簡単な装置構成で短時間で物体の三次元的な位置が把握できる画像処理装置およびこの画像処理装置を装備した物体移載装置を提供することを目的とする。

【0017】

【課題を解決するための手段】この発明の請求項1に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から積載された物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段とによって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものである。

【0018】この発明の請求項2に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から積載された物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出

手段と、認識対象物体の寸法データを格納する物体寸法データベースと、物体の寸法データをもとに物体の2次元画像上での標準的パターン画像を生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段とによって得られた個々の物体の距離情報とを統合する情報統合手段とを備えた画像処理装置と物体を把持し移載する物体移載手段とを備えたものである。

【0019】この発明の請求項3に係る画像処理装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行なうステレオ画像ブロック対応付手段で構成されたものである。

【0020】この発明の請求項4に係る物体移載装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行なうステレオ画像ブロック対応付手段で構成された画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0021】この発明の請求項5に係る画像処理装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されているものである。

【0022】この発明の請求項6に係る物体移載装置は、距離画像生成手段が解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成された画像処理装置と、物体を把持し移載する物体移載手段とを備えたものである。

【0023】この発明の請求項7に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列挙生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと

スト、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものである。

【0024】この発明の請求項8に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列挙生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行い、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置と、物体を把持し移載する移載手段を備えたものである。

【0025】この発明の請求項9に係る画像処理装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有するものである。

【0026】この発明の請求項10に係る物体移載装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と、物体を把持し移載する移載手段とを備えたものである。

【0027】

【発明の実施の形態】実施の形態1. 実施の形態1の構成を示すブロック図を図1、動作の流れを示すフローチャートを図2、生成された各段階の画像を図3に示す。ここでは動作説明が理解され易いように認識すべき物体は、例えば段ボール箱のような多面体形状を想定し、物体は同一形状および同一サイズのもが積まれている状態として説明する。

【0028】図1において、11は所定の間隔において配置された2台のカメラで構成された画像入力手段、1

2は画像入力手段11の撮像した画像から距離画像を生成する距離画像生成手段、13は距離画像から最上段面を抽出する最上段面領域抽出手段、14は最上段物体候補抽出手段、15は原画像を格納する濃淡画像格納手段、16は物体寸法データベース、17は物体寸法データベース16に格納された情報を基に2次元パターンとしての基準パターンを自動的に生成する2次元基準パターン生成手段、18は2次元的な物体位置を検出する物体位置検出手段、19は物体位置検出手段18によって検出された物体位置と距離画像の距離情報を統合する情報統合手段である。

【0029】図3の(a)は画像入力手段11が撮像した原画像、(b)は最上段面領域抽出手段13が抽出した最上段物体領域、(c)は2次元基準パターン生成手段により物体寸法データを用いて生成された基準パターンのテンプレート、(d)は最上段物体領域から抽出された候補の一つを示す模式図、(e)はテンプレートマッチング法による位置決め状況を示す。

【0030】ステップST101で、装置が起動されると、画像入力手段11によって図3(a)に示す認識対象の画像が距離画像生成手段12に入力される。ステップST102では、距離画像生成手段12によって距離画像が生成される。距離画像の生成手段としては区間コード化法またはステレオ対応探索法等がある。ステップST103で最上段物体領域抽出手段13により、図3(b)に示す最上段の物体に相当する高さをもつ最上段物体領域が抽出される。ステップST104で画像入力手段11で撮像した原画像が濃淡画像格納手段15に格納され、ステップST105では2次元基準パターン生成手段17によって、物体寸法データベース16に格納された情報をもとに2次元パターンとしての基準パターンが自動的に生成される。この基準パターンはテンプレートと呼ばれ、図3(c)に示すようにテンプレートは物体の輪郭部分を表現した輪郭プレートとなっている。

【0031】ステップST106では、2値で表現された最上段物体領域の物体は画像上では矩形パターンに見えることから、直角に交わる辺縁部を検出することによって候補を抽出される。上記の通り距離画像の解像度が粗いことから、候補の位置は不正確である。ステップST107において、前記物体候補が物体位置検出手段17によって位置決めされ、位置決めには、前記2次元基準パターンが用いられ、テンプレートマッチング法による処理がなされる。ただし、大まかな位置は既に候補抽出の段階で検出されていることから、マッチングは候補検出位置からそのごく近傍のみを探索する処理で十分である。したがってテンプレートマッチングも高速に実行可能である。位置決めの様子は図3(e)に示す。物体位置が検出されたならば、その2次元的な物体位置情報は、情報統合手段19によって前記距離画像生成手段12による距離情報と統合され、それが最終的な結果とし



て出力される。

【0032】ステップST108では、すべての物体を検出したかどうかを判定し、もし検出していなければ処理流れのうちステップST106から再び実行され、別の物体候補について一連の処理ステップST106、ST107が実施される。もし、すべての物体が検出されていれば、ステップST109で認識処理を終了させる。以上の通り、この発明により粗く距離画像を生成し、その結果を元に物体の候補を抽出して詳細な位置検出は2次元の汎用的な処理手法であるテンプレートマッチングをそのパターン探索の範囲を非常に限定した形で実行することで、距離画像生成装置の装置規模を小さくすることができるものである。

【0033】なお、距離画像生成手段として、空間コード化法を解像度を粗くして用いる方法、あるいは、ステレオ視覚やスリット光走査による光切断法のような方法であってもよく、粗い解像度の距離画像が生成され得る装置であればよい。

【0034】また、輪郭情報によるテンプレートマッチングによる位置決め例を示したが、同じく物体の輪郭を直線の組合せと考えると直線として辺ごとに位置合わせを行う方法を用いてもよい。また、物体寸法データベースの内容を使い、通常の2次元テンプレートマッチングを用いても良い。また、テンプレートマッチング処理内部で使用されるデータ格納形式に関して特に明示しなかったが、2次元画像として格納する方法他に、輪郭部分のエッジの位置を1次的に内部に羅列して格納する方法を用いてもよい。

【0035】さらに、最上段物体の候補を抽出する際に、距離画像から得られた最上段面領域情報からそのコーナー部分を利用する手法を示したが、この他にも例えば上記の2値の距離画像を用いて、別途用意した2値の距離画像としてのテンプレートを使って2値のテンプレートマッチングの方法によって粗く位置を検出する方法を適用しても同様の効果を奏する。

【0036】図2の動作フローチャートは、一例であり、図1に示した各手段の入出力関係が適正でさえあれば別の処理流れであってもよい。例えば、距離画像生成と濃淡原画像入力の手ステップはどちらが先に行なわれてもよい。

【0037】以上は、認識すべき対象物体として多面体即ち箱状の物体の単一品種物体を想定した説明であったが、例えばセメント袋、米袋などのように袋状の物体であっても、この発明の構成を何ら変えることなく同一の構成で画像認識を実現することができる。次に袋状物体の認識する場合の状況について、上記箱状物体における場合との違いを説明する。

【0038】図2のステップST105において、2次元基準パターン生成手段16によって2次元基準パターンが自動的に生成される。これは物体の位置決めを行な

うためのテンプレートマッチングにおけるテンプレート生成に相当する。袋状物体の場合は辺縁が変形するなど箱状物体とは異なり形状が不安定であるので、前記基準パターンとしては図3のような矩形の輪郭部分を取り出しただけでは不十分である。この場合は物体のコーナー部のみを輪郭で表現した部分テンプレート、あるいは物体の辺部を表現した直線輪郭部分テンプレートを自動的に生成する。いずれの場合も、袋状物体の寸法を格納した物体寸法データベース15の内容を参照して生成される。

【0039】ステップST106の物体候補抽出においては、袋状物体の場合は上で述べたような直角に交わる縁部などの情報を用いて候補を抽出することは困難である。この場合は最上段物体領域が表現された画像を2値画像とみなし、物体寸法データベース15に格納されている袋状物体モデルに関する情報から2値の物体テンプレートを自動的に生成し、これらを粗く2値テンプレートマッチングさせることで大まかな候補抽出が可能となる。

【0040】ステップST107では、物体の位置が検出されるが、ここでは前記のような物体のコーナー部分、あるいは、直線辺縁部分の輪郭テンプレートが部分テンプレートとして入力画像から得たエッジ画像とマッチングされる。物体の位置はマッチングの結果として、複数の部分テンプレートのマッチングした位置から総合的に判断、計算される、以上のように、袋状の物体に対しても箱状物体に対する認識と同じ程度に動作させることができる。

【0041】また、以上説明した実施の形態1では、単一の品種の物体が積まれている状況を想定したが、物体寸法データベース16に複数の物体の寸法データを格納しておき、複数の2次元基準パターンを自動的に生成し、物体位置検出手段17においては前記複数の基準パターンを全てマッチングさせて最も類似性の高いテンプレートが存在する物体であるとするような動作をさせることで、異なる種類の物体が複数個積まれている状況に対しても適確に動作させることができる。

【0042】上記構成の画像処理装置を物体移載装置に組み込むことにより、高精度のパターン投光器、大容量の画像メモリを必要とせず、また画像入力時間も短時間で実行する正確に動作する物体移載装置となる。

【0043】実施の形態2. 実施の形態2の構成を示すブロック図を図4、動作の流れを示すフローチャートを図5に示す。図6はランダムドットパターン照射の説明図、図7はステレオ画像からブロック対応付処理の方法を示す説明図である。この実施の形態2. は、実施の形態1の距離画像生成手段の部分ランダムテクスチャパターン投光によりステレオ対応付して距離画像を生成するようにしたものであり、この点が相違し、その他の構成は実施の形態1と同じである。

【0044】図4において、20はランダムドットパターンを投光するランダムテクスチャパターン投光手段、21は画像入力手段であり、第一画像入力手段21aと第二の画像入力手段21bとで構成されている。22はステレオ画像ブロック対応付手段である。その他の最上段画像領域抽出手段13、最上段物体候補抽出手段14、濃淡原画像格納手段15、物体寸法データベース16、2次元基準パターン生成手段17、物体位置検出手段18および情報統合手段19は実施の形態1. と同一の構成である。

【0045】以下この実施の形態2について図5のフローチャートにしたがって動作を説明する。ステップST201で電源がONされると、ランダムテクスチャパターン投光手段20の電源が入り、認識対象物に対してランダムドットパターンが投影される。ランダムテクスチャパターン投光手段は図6(a)に示すように、パターン投光器30により、認識すべき対象物に対してランダムドットパターンを投光するものである。パターン投光器は図6のように物体の上方に下向きに設置し、図6

(b)に示したようなランダムドットを認識物体に投影する。実施の形態2. では、認識すべき対象物体は、段ボール箱であるが、例えばセメント袋のような袋状の物体であっても認識可能である。対象物の上面にランダムドットが投影され、ステップST203では、第一の画像入力手段21aおよび第二の画像入力手段21bによって撮像された一対の画像のステレオ画像が入力される。このステレオ画像は、光軸を所定の距離はなして同一の解像度、同一の焦点距離の二つのカメラで撮像される。

【0046】次にランダムドットパターンについて説明する。図6(b)に示したランダムドットは、この実施の形態2では、128×128の解像度で構成され、各画素は正方形であり、計算機によって一様乱数的に各画素ごとに白か黒かの値が割り当てられ、白と黒の画素数の比率はほぼ1:1となっている。パターン投光器30

$$S_{SAB}(d_x, d_y) = \sum_{j=1}^N \sum_{i=1}^N |L(d_x + i, d_y + j) - b_R(i, j)| \cdots \cdots \cdots \text{(式1)}$$

【0049】(式1)で求めた $S_{SAB}$ が最小になる( $d_x, d_y$ )を検出することで、注目ブロックbRに対応する左画像上のブロックbLが見つかったことになる。ステレオ法は、このようにして得たブロック間の位置から、その差を視差とし、三角測量の原理によってカメラから物体までの距離を得る方法である。以上説明した一連のブロック対応付処理を右画像上のすべての小ブロックに対して繰り返して行い、各ブロックごとに対応する左画像上の位置を探索し、距離画像を生成する。

【0050】これ以降は実施の形態1と同じである。ステップST206において、最上段面領域抽出手段13によって距離画像から最上段物体面に相当する領域が抽

には投影パターンの焦点を調節する機構が設けられており、投光器と対象物体の大まかな距離がわかっている場合にはそのデータをもとに物体表面で合焦するように焦点が調節される。あらかじめ全く距離がわかっていない場合には、事前に調整された距離において焦点が合うように動作する。

【0047】ステレオ画像ペアの入力が終わると、ステップST204でランダムテクスチャパターン投光手段20がOFFされる。次にステップST205でステレオ画像がブロック対応付手段22によってステレオ画像が処理される。図7(a)および(b)は対象物体を撮影したステレオ画像の一対の画像である。図7(a)を左画像、図7(b)を右画像と呼ぶこととする。ブロックマッチングは右画像を図のように格子状の小ブロックに分割することから行なう。いま、分割されて生成された複数の小ブロックのうち、注目しているブロックbRを注目ブロックと呼ぶ、実際には左および右画像には上記ランダムドットパターンが投影されているが、図を簡略化するためにここではパターンは描いていない。この注目ブロックbRに対応する左画像上での位置を探索する。別途設定した探索範囲の中でbRと同じパターンを有するパターンを探索する。探索の方法として累積差分絶対値最小化による探索を採用する。これは左画像上の探索範囲の中を注目ブロックbRを動かしながら逐次重ね合わせ、その類似性をもっとも良くなる位置をもって対応点とするものである。この際、類似性の評価としては、累積差分絶対値を用いる。例えば、注目ブロックをbR( $i, j$ )、左画像をbL( $i, j$ )、ブロックサイズを $N \times N$ とすると、左画像上の位置( $d_x, d_y$ )における累積差分絶対値 $S_{SAB}(d_x, d_y)$ は、(式1)で表される。この実施の形態2. では $N=16$ である。

【0048】

【数1】

出され、続くステップST207では、第一の画像入力手段20a、または第二の画像入力手段20bによって撮像された濃淡原画像が濃淡原画像格納手段15に格納される。この時は当然のことながらランダムドットパターンは投影されていない。ステップST208では、物体寸法データベース16に格納されている認識対象物体の寸法データを使って、2次元基準パターン生成手段17によって2次元基準パターンが自動的に生成される。

【0051】ステップST209では、上記最上段面領域が抽出された画像は、最上段面とそれ以外の面の2つに分けられ、2値化され、最上段物体候補抽出手段14によって最上段物体候補の一つが抽出される。ステップ

ST210では、抽出された一つの物体候補の位置近傍に関し、上記2次元基準パターンと物体位置検出手段15に格納された濃淡原画像情報を用いて、物体位置検出手段18によって候補物体の精密な位置が計測され検出される。ステップ211で、決定した候補物体の2次元位置は、情報統合手段19によって前記距離画像情報と統合され、ステップST212で全ての物体を認識したと判断するまで、ステップST209からST211までの処理が繰り返され、ステップST212の判断が全ての物体が認識したと判断されれば、最終的に物体の3次元情報として出力される。

【0052】ステップST209からST211までの処理によってすべての物体を認識したと判断されればステップST213で処理を終了する。すべての物体が認識されたかどうかの判定は、上記最上段面領域が抽出された2値化された距離画像をもとに順次物体の位置が確定したものを消去し、なおかつ画像上に残っている最上段面領域が存在するか否かを調べることによって行う。

【0053】以上説明したように、この実施の形態2では、認識対象物体に対してランダムドットパターンを投光するので、例えば表面に図柄の存在しない無地の段ボール箱であっても、その表面各位置における距離データを得ることが可能であり、物体表面に宛先ラベルなどの位置不定なテクスチャに影響されないステレオ画像を得ることが可能である。またステレオ対応づけのためのマッチング処理も単純なブロックマッチングを適用することができる。

【0054】なお、この実施の形態2ではカメラの解像度として、カメラ視野に対して128×128画素のランダムドットパターンを投影したが、認識対象のサイズによってはさらに細かい、または粗いドットパターンを投影しても良く、またブロックマッチングのブロックサイズとして16×16画素のサイズを採用したが、8×8画素、32×32画素等でもよい。また、ブロックの形状も必ずしも正方形でなくても良い。さらにマッチング時の探索範囲は、図7では長方形に設定したが、状況に応じてさまざまに変化させることでマッチングの信頼性を向上させたり処理時間を短縮することもできることは一般のステレオ対応と同様である。

【0055】さらに、この実施の形態2では、ランダムドットパターンとして白黒の2値正方形ドットを用いたが、例えば濃淡をつけたドットパターンや着色したパターンであっても良い。また形状についてもさまざまな大きさのさまざまな形状のマークを組み合わせ使用しても良く、また、パターン投光器と対象物体との幾何学的位置関係については、物体に対して鉛直上方から真下に向けてパターンを投影する例を示したが、斜めから投影しても同様に距離画像が検出できる。

【0056】また投光するパターンとしてランダムな位置に発生させたドットパターンを使った例を示したが、

ブロック対応づけ処理において対応点を探索する場合にほとんど水平方向の探索になることを考えると、例えば周期のランダムな縦スリット群、あるいは幅のランダムな縦スリット群をパターンとして投影しても同様の効果を奏することはいうまでもない。

【0057】また、パターン投光器の構成方法に関しては、特に詳しく説明しなかったが、ランダムドットの投影位置精度は高精度が不必要であり、処理の途中で投影パターンを変化させる必要がないことから、例えば従来技術で示したように、液晶プロジェクタのような複雑な投影装置は不要であり、たとえば通常の家庭用のスライド上映機のような簡単な構成の装置で十分である。このことは本発明の装置構成上の利点の一つである。

【0058】実施の形態2は、既に述べたように、認識対象物体として段ボール箱、即ち手面体状の物体を想定して動作を説明したが、その他の形状の物体、例えばセメント袋のような袋状の物体であっても本発明の構成により容易に距離画像を生成することができ、前記本実施例における認識手順により対象物の3次元的位置を計測することができる。なお、距離画像生成より以降の処理については前述の実施の形態1と同じである。即ち2次元基準パターンとして物体のコーナー部分や直線輪郭部をもち、それぞれの認識対象との位置決め結果を統合することで最終的な物体位置検出が可能である。このように実施の形態2は物体の形状が多面体の場合だけでなく、袋のような形状その他の場合でも物体の3次元情報を認識することができる。

【0059】実施の形態3、実施の形態3の構成を示すブロック図を図8、動作の流れを示すフローチャートを図9に示す。図10は低解像度距離画像から段差領域部を抽出し、高解像度距離画像の段差領域部のみ高解像度のステレオ対応付を行う状況の説明図である。図8において、13～19は、実施の形態2、と同一機能を有するものであり説明は省略する。31は画像入力手段21が低解像度で撮像したステレオ画像から低解像度の距離画像を生成する低解像度ブロック対応付手段、32は低解像度で段差領域を抽出する段差領域抽出手段、33は低解像度の画像から抽出した段差領域部について高解像度の距離画像を生成する高解像度ブロック対応付手段、34は低解像度距離画像と、この距離画像から抽出した段差領域部の高解像度距離画像とを合成する距離画像合成手段である。

【0060】ステップST301で電源がONされ起動すると、ステップST302で、ランダムテクスチャパターン投光手段20のランプ電源がONされ、認識対象物に対してランダムドットパターンが投影される。ステップST303では、一対のカメラで構成される画像入力手段21により一対の画像のステレオ画像が低解像度ブロック対応付手段31および高解像度ブロック対応付手段32に入力される。ステレオ画像が入力が終わると



ステップST304でランダムテクスチャパターン投光手段20はOFFする。ステップST305では第一の画像入力手段21aまたは第二の画像入力手段21bから対象物体の濃淡画像が濃淡画像格納手段に15に格納される。

【0061】ステップST306では、低解像度ブロック対応づけ手段31によってステレオ画像が処理される。図10(a)は低解像度ブロック対応付手段31によって得られた低解像度距離画像部分Aを模式的に表した図である。画像入力手段として512×512画素の解像度のCCDカメラを使用しているが、低解像度距離画像としては32×32画素の距離画像を生成し、ブロックマッチングの方法は実施の形態1で示した方法と同様の方法で行い、例えばブロックサイズを16×16画素にすることで得られる。ステップST307では上記低解像度距離画像をもとに、段差領域抽出手段32で対象物の段差領域部分Bを抽出する。段差領域部分Bの模式図を図10(b)に示す。ステップST308では、高解像度ブロック対応付手段33により、上記段差領域部分(B-Aの斜線部分)についてのみ、より高解像度の距離画像を生成する。この場合は高解像度とは128×128画素である。この処理はブロックマッチング処理においてブロックサイズを8×8に設定し、かつ、ブロックを4画素づつずらしながら対応点を検出することによって得られ、図10(c)に段差領域部分の距離画像Cを示す。図10(c)において、斜線部分(Bの枠内)は低解像度距離画像から得られた物体の段差領域、灰色の部分(Cの枠内)は、高解像度距離画像生成の結果新たに得られた距離画像の段差領域部分に相当する部分である。ステップST309では、距離画像合成手段34によって上記低解像度距離画像と高解像度距離画像を合成し、図10(d)のような128×128の解像度の距離画像が得られる。

【0062】ステップST310以降は、実施の形態2と同様に処理される。ステップ310では、最上段面領域抽出手段13により、距離画像から最上段物体面に相当する最上段領域が抽出され、ステップST311では、物体寸法データベース16に格納されている認識対象物体の寸法データを使って、2次元基準パターン生成手段17によって2次元基準パターンが自動的に生成される。ステップST312では、上記最上段面領域が抽出された画像が、最上段面とそれ以外の面の2つに分けられて2値化され、最上段物体候補抽出手段14によって最上段物体候補の一つが抽出される。ステップST313では、抽出された一つの物体候補の位置近傍に関し、上記2次元基準パターンと、上記濃淡原画像情報を用いて、物体位置検出手段18によって候補物体の精密な位置が計測され検出される。このようにして決定した候補物体の2次元位置は、情報統合手段19によって上記合成された高解像度の距離画像情報と統合され、最終

的に物体の3次元情報として出力する。ステップST314では、ステップST312、ST313の処理によって、すべての物体を認識したか、否かの判断を行い認識したと判断されればステップST315で処理を終了となり、認識していない物体がまだ存在すると判断されれば、ステップST312、ST313に戻って物体候補を抽出する処理を繰り返す。すべての物体が認識されたかどうかの判定は、上記最上段面領域が抽出された2値化された距離画像をもとに順次物体の位置が確定したものを消去し、なおかつ画像上に残っている最上段面領域が存在するかどうか調べることによって行う。

【0063】以上のように、この実施の形態3は、最終的に高解像度の距離画像を得るために、予め低解像度の距離画像を短時間で得ておき、物体の段差部分に相当する領域に限定して高解像度のステレオ対応付処理を行なうことにより、少ない演算時間で高い解像度の距離画像を生成することを可能としたものであり、物体候補の抽出精度等も向上させることができる。

【0064】低解像度距離画像は32×32画素、高解像度距離画像は128×128画素として使用したが、許容される処理時間や認識したい対象物体の寸法などを考慮して、これと異なる解像度を設定してもよく、高速処理の低解像度距離画像生成の時間と、高解像度ステレオ対応付処理の組合せによるメリットは失われることはない。

【0065】また、低解像度距離画像生成時のステレオ画像と、高解像度距離画像生成時のステレオ画像を同じ画像として構成したが、低解像度ステレオ画像として、フル解像度(CCDカメラ解像度と同じ解像度)の画像を1/2や1/4などに縮退させて得た縮退画像をもとにブロックマッチングを行なってもよい。さらに、高解像度距離画像生成時には第一および第二の画像入力手段21のCCDカメラのレンズをズームアップさせ、低解像度距離画像における物体の段差部分をより詳細に撮像するようにしても良い。

【0066】さらに、対象物体に投影するランダムドットパターンを一定のパターンとしたが、2種類の異なる大きさのドットパターン投光器を用意し、低解像度距離画像生成時には大きなドットパターン投影を行ない、高解像度距離画像生成時にはより小さなドットパターン投影を行なうように変形して実施すれば、精度の高い距離画像が得られる。

【0067】実施の形態4. 実施の形態4の構成を示すブロック図を図11、動作の流れを示すフローチャートを図12に示す。図11において、11～13、15～17、19、21は実施の形態1と同一機能を有するものであり説明は省略する。図において、64は物体候補組合せ仮説生成手段、65は仮説妥当性検証手段、69は物体位置検出手段である。

【0068】ステップST601で、装置が起動される

と、画像入力手段11の2つのカメラで対象物体の画像が距離画像生成手段12に入力され、個別に格納される。2つのカメラは、光軸を共有しない同一の仕様のものが使用される。ステップST602で、距離画像生成手段12によって上記2枚のステレオ画像がステレオマッチングされ、その視差から距離画像が生成される。

【0069】ステップST603で、画像入力手段11により撮像された濃淡原画像が濃淡原画像格納手段に格納される。この実施の形態6.の認識対象とする物体の濃淡原画像を図13(a)に示す。物体は段ボール箱等の箱状の物体であり、それらが何段かに積載されている。ステップST604で距離画像はから、最上段面領域抽出手段13によって、最上段物体の上面部分の画像から領域として抽出される。図13(b)は、抽出された最上段物体領域である。図のように、最上段以外の領域も示されている。

【0070】ステップST605では、箱状物体の寸法データを格納した物体寸法データベース16に格納されたデータを用いて、2次元基準パターン生成手段17によって物体の2次元的な画像パターンを表現した基準パターンを自動的に生成する。図13(c)にその基準パターンの模式図を示す。

【0071】以下の動作の説明が、この実施の形態4の本質となる部分である。ステップST607において、物候補組合せ仮説生成手段64によって、上記抽出された最上段面領域と2次元基準パターンとから、物体候補の組合せ仮説が一つ生成される。仮説とは図13(c)に示されたようなものであり、この実施の形態4では、この仮説を「仮説1」と呼ぶ。仮説は図13(b)のような最上段物体領域を画像で表現したものに対し、認識対象物の基準パターンをパターンマッチング、例えばテンプレートマッチングの技法によって処理し、図13

(b)のような最上段面領域を形成する2次元基準パターンの組合せを類推することによって得られる。また、同時に、物体位置検出手段69によって図13(d)のように物体2次元的位置が検出される。物体位置検出は、上記格納された濃淡原画像と上記2次元基準パターンとの比較によって行う。2次元基準パターンの輪郭部分の情報を用い、上記濃淡画像から検出したエッジ画像との間でテンプレートマッチングを行なう。その際、仮説生成の段階で物体の粗い位置は既知なので、テンプレートマッチング処理におけるパターン探索領域は上記位置近傍のみでよい。

【0072】ステップST608では、上記仮説が妥当であるか否かが検証される。仮説妥当性検証手段65は図13(d)のような仮説画像を生成し、物体候補1から6までの6個の物体の組合せによって、図13(b)のような最上段面領域が合理的に生成され得るかどうかを計算する。即ち、上記6個の物体候補によって生成される仮説画像と、最上段面画像とを比較し、その差とな

る領域の面積を算出して上記面積が大きいほど誤った仮説であると判定するように動作する。ステップST608ではこのようにステップST606で生成した一つの仮説に対し、その妥当性指標を一つ与える。上記面積の逆数が妥当性指標となる。

【0073】ステップST609では、ステップST606からST608までの処理を考えられる仮説をすべて生成するまで繰り返し、ステップST610では、仮説をすべて生成し、それらの中でもっとも妥当性の高い仮説を選択する。即ち、各仮説についての前記妥当性指標を比較し、最も高い指標を持つ仮説を選ぶ。図13

(e)(f)は、それぞれ別の仮説2および仮説3である。図から明らかなように、この場合は仮説2が最も高い妥当性指標を持つのでこれが最もよい仮説として、ステップST610において選択されることになる。

【0074】ステップST611では、情報統合手段19によって、上記の最もよい仮説に相当する2次元位置情報と前記距離画像が統合され、ステップST612で認識結果として装置から出力されて処理を終了する。

【0075】この実施の形態4は、距離画像生成手段として2つのカメラによるステレオ視を用いたが、3つ以上のカメラによるステレオ視、ランダムドット等の特徴ある投光パターンの投光を併用したステレオ視、あるいは従来技術として説明した空間コード化法による距離画像生成手段でもよい。この実施の形態4は、距離画像から得た認識結果の仮説を検証して尤もらしい結果を得ることに特徴があるので、距離画像の生成方法はどんなものでも良い。

【0076】また、この実施の形態4は、仮説の妥当性を検証する方法として、仮説画像と実際の画像との差をもとに妥当性指標を計算する方法を用いたが、例えば、上記仮説を得る際に計算されたテンプレートマッチングの類似度のすべての候補物体に対する総和が大きいほど妥当性が高いとするような指標を用いても同様の効果を得ることができる。

【0077】また、この実施の形態4では物体位置の検出方法として、基準パターンから得た物体の輪郭情報と、濃淡原画像から生成したエッジ画像とをテンプレートマッチング手法により処理していたが、これを2値画像のテンプレートマッチングなどの位置決め手法を用いても同様の効果を得ることができる。

【0078】さらに、この実施の形態4では、物体候補の組合せ仮説の生成と検証を逐次行なうような処理フローを説明したが、多数の仮説を一度に生成して格納しておき、それらの妥当性検証を一度に行って最も妥当性の高いものを選択するようにしてもよい。

【0079】さらに、実施の形態4では、すべての仮説を生成するように説明したが、仮説は物体候補の組合せとして得られることから、物体の組合せ最適化問題として扱い、近似解を求めても同様の効果が得られる。

【0080】実施の形態5。実施の形態5の構成を示すブロック図を図14、動作の流れを示すフローチャートを図15に示す。図14において、11～13、15～19は実施の形態6。と同一機能を有するものであり説明は省略する。71は物体配列基準データベース、72は物体配列判定手段、73は警告発生手段である。

【0081】この実施の形態5は、実施の形態4の構成に、物体配列が誤って配列されているときに警告を発生するように構成したものである。したがって、図15のフローのステップST701～ST711は、図12のステップ601～611と同じである。

【0082】ステップST701で装置が起動されると、画像入力手段11によって対象物体の画像が距離画像生成手段に入力され、個別に格納される。ステップST702で、距離画像生成手段12によって上記2枚の画像がステレオマッチングされ、その視差から距離画像が生成される。この処理は、必ずしも単純なステレオマッチングである必要性はなく、例えば実施の形態2.に示したようなパターン投光器を併用したステレオ視覚であってもよい。

【0083】ステップST703では、画像入力手段11により認識対象物の濃淡原画像が入力され、濃淡原画像格納手段15に格納される。ステップST704で最上段面領域抽出手段13により最上段面領域が抽出される。ステップST705では、箱状物体の寸法データを格納した物体寸法データベース16を用いて2次元基準パターン生成手段17によって物体の2次元的な画像パターンを表現した基準パターンを自動的に生成する。ステップST706では、物体候補組合せ仮説生成手段64によって、上記抽出された最上段面領域と基準パターンとから、物体候補の組合せ仮説が一つ生成される。仮説とは実施の形態6.で説明したと同様の図13(d)に示されたようなものである。この実施の形態7.においては、仮説は図13(b)のような最上段物体領域を画像で表現したものに対し、認識対象物の基準パターンをパターンマッチング、例えばテンプレートマッチングの技法によって処理し、図13(b)のような最上段面領域を形成する2次元基準パターンの組合せを類推することによって得ている。ステップST707では、物体位置検出手段69によって図13(d)のように上記仮説に含まれる候補物体それぞれの2次元的位置が検出される。物体位置検出は、上記格納された濃淡原画像と、上記2次元基準パターンの輪郭部分の情報を用い、濃淡画像から検出したエッジ画像との間でテンプレートマッチングによって検出される。

【0084】ステップST708では、上記複数の物体候補によって生成される仮説画像と、最上段面画像とを比較し、その差となる領域の面積を算出して大きいほど誤った仮説である可能性が高いと判定する。ステップ709では、ステップST706～708の処理を考えら

れる仮説をすべて生成するまで繰り返し、ステップ710では、仮説をすべて生成し、それらの中でもっとも妥当性の高い仮説を選択する。即ち、各仮説についての前記妥当性指標を比較し、最も高い指標を持つ仮説を選ぶ。

【0085】ステップST711では、情報統合手段19によって上記仮説で最もよい仮説に相当する2次元位置情報と上記距離画像が統合される。この時点で、この実施の形態5における画像処理装置は物体の配列および各々の物体の3次元的位置を知ったことになり装置内部に格納される。ステップ712では、物体配列判定手段72によって上記格納された認識結果が、予め用意された正しい配列データと比べて合致しているか否かが判定される。正しい配列データは、物体配列基準データベース71に格納された被認識物体によって決められた配列パターンである。例えば、この実施の形態5による装置が使用されるような物流工場においては、認識対象物体としての段ボール箱はその形状や寸法によって予め正しい「積みパターン」と呼ばれるパレット上への箱の積み方が決められている。

【0086】上記物体配列基準データベースにはそのような基準とすべき物体の配列が記述されている。ステップST712において認識した結果が基準となる積みパターンと合致していない場合には、ステップST713において、警告発生手段73によって装置使用者や上位システムの管理者に対して警告を発生して物体が正しく積まれていないことを知らせる。この実施の形態5は、ブザーによる音と、装置に接続された表示用ディスプレイ装置への文字表示によって警告が発生される。ステップST712において認識結果が正しいパターンであると判定された場合は、情報統合手段19によって最終的な物体認識結果が出力され、例えばロボットにデータが送信されることによりロボットによる物体の把持動作が行なわれることになる。認識動作としてはステップST715で動作を終了する。

【0087】この実施の形態5では、このように物体を認識するだけでなく、その配列パターンを基準データと比較することによって予想された配列なのかどうかを判定し、さらに上記認識結果が基準データと異なる場合には警告を発生する機能を有している。これにより、作業者が自分の積み作業ミスに気づいたり輸送中の激しい荷崩れを検知できるという利点がある。

【0088】なお、距離画像生成手段として2つのカメラによるステレオ視を、仮説の妥当性検証の方法として仮説画像と実際の画像との差をもとに検証する方法を、そして物体位置の検出方法として物体の輪郭情報を利用したテンプレートマッチング手法をそれぞれ利用している。しかし、これら部分的な処理手段処理手段は、他の手法により実現しても同様の効果を奏することはいうまでもない。例えば、距離画像生成手段として、実施の形

態2. で説明したパターンを併用するステレオ視を、仮説の妥当性検証方法として各物体を濃淡基準パターンと比較した類似度の総和が高いほど妥当性が高いと判断されるような方法を物体位置検出方法として仮説に含まれる各物体を予め用意した濃淡画像パターンと正規化相互相関によるテンプレートマッチング法を、それぞれ採用してもよい。

【0089】また、認識結果が予め格納された物体配列基準データベースと比較して合わないと判断した時に警告を発生する例を説明したが、基準データとの類似度合に応じて、警告の内容を変えて装置使用者や作業者に知らしめるようにしてもよい。例えば、制御表示器に『自動認識の結果判定された積み方の正しさはレベル8です』などと表示する。さらに認識した結果、積荷の荷崩れ状態が非常に顕著であり、このまま作業を継続することが適当でないと判断される場合には上記警告を発生するだけでなく、上位の管理計算機にシステムの停止を要求したり、または本装置が直接システム停止したりするようにしてもよい。

【0090】さらに、警告手段としてブザーと表示器による文字表示を採用したが、この他にも人間に警告を送る手段として、合成された人間の声を発生させるなどの方法を用いてもよい。

【0091】

【発明の効果】この発明の請求項1に係る画像処理装置は、複数の物体の距離画像を生成する距離画像生成手段と、距離画像から物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データをもとに物体の2次元画像上での2次元基準パターンを生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものとしたことにより、認識対象物体が粗い距離画像と2次元基準パターンのマッチングによって物体候補が検出できるので、画像メモリは小さな容量のものでよく、画像入力時間が短くなり、個々の物体が三次元的に明確に検出できる。

【0092】この発明の請求項2に係る物体移載装置は、複数の物体の距離画像を生成する距離画像生成手段と、距離画像から物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域から物体を個々に分離して抽出する最上段物体候補抽出手段と、認識対象物体の寸法データをもとに物体の2次元画像上での2次元基準パターンを生成する2次元基準パターン生成手段と、カメラから入力された濃淡原画像を格納する濃淡原画像

格納手段と、個別に抽出された最上段物体候補について2次元基準パターンと濃淡原画像の情報をを用いて物体の位置を検出する物体位置検出手段と、各手段で得られた物体位置の検出結果と距離画像生成手段とによって得られた個々の物体の距離情報とを統合する情報統合手段とを備えた画像処理装置を物体移載装置に組み合わせたものであり、このようにしたことにより、物体移載装置として、物体の検出時間が短く、位置は正確に検出され、物体の移載は素早く正確に行うことができる。

【0093】この発明の請求項3に係る画像処理装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行なうステレオ画像ブロック対応付手段とで構成したので、認識対象物体に対してランダムドットパターンが投光されるため、表面に図柄のない無地の物体、あるいは表面に光が反射するラベルが貼ってあっても、表面各位値のにおける距離データを正確に得ることができる効果がある。

【0094】この発明の請求項4に係る物体移載装置は、距離画像生成手段がランダムなテクスチャパターンを投光するランダムテクスチャ投光手段と、ステレオ画像を入力する第一の画像入力手段および第二の画像入力手段と、第一および第二の画像入力手段が撮像した二つの画像間の対応付を行なうステレオ画像ブロック対応付手段で構成された画像処理装置を物体移載手段に組み合わせたものであり、このようにしたことにより、認識対象物体に対してランダムドットパターンが投光されるため、表面に図柄のない無地の物体、あるいは表面に光が反射するラベルが貼ってあっても、表面各位値が正確に得られ、物体移載装置として、物体の検出時間が短く、物体の位置は正確に検出され、物体の移載は素早く正確に行うことができる。

【0095】この発明の請求項5に係る画像処理装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成されたものとしたので、少ない画像メモリで、短い演算時間で、高解像度の距離画像が生成され、物体候補の抽出精度も高くすることができる効果がある。

【0096】この発明の請求項6に係る物体移載装置は、距離画像生成手段が、解像度の粗い距離画像を生成する低解像度距離画像生成手段と、低解像度距離画像から物体の段差領域部を抽出する段差領域抽出手段と、段差領域に対して高い解像度の距離画像を生成する高解像

度距離画像生成手段と、低解像度距離画像と高解像度距離画像の解像度の異なる2つの距離画像を合成する距離画像合成手段とで構成された画像処理装置を、物体移載手段に組み合わせたものであり、短い演算時間で、高解像度の距離画像が得られ、物体候補の抽出時間が短く、抽出精度が高く、物体の位置は正確に検出され物体の移載は素早く正確に行うことができる。

【0097】この発明の請求項7に係る画像処理装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列挙生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備えたものとし、距離画像から抽出された最上断面領域に対して、対象物体の配置仮説から候補物体の位置を検出するので、不規則に積載された対象物体においても正確に位置を検出することができる。

【0098】この発明の請求項8に係る物体移載装置は、積載された複数の物体の距離画像を生成する距離画像生成手段と、距離画像から最上段に位置する物体の最上段面領域を抽出する最上段面領域抽出手段と、最上段面領域抽出手段の出力から物体候補の組合せを仮説として複数列挙生成する物体候補組合せ仮説生成手段と、認識対象物体の寸法データを格納する物体寸法データベースと、2次元基準パターンと濃淡原画像に含まれる情報を使って上記仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段と、複数の仮説のうち最も評価値の高い仮説をもとに認識を行ない、この認識した画像と距離画像生成手段によって得られた個々の物体の距離情報とを統合する情報統合手段とを備え、個々の対象物体の三次元位置情報を出力する画像処理装置と物体移載手段とを組み合わせたものとしたので、不規則に積載された対象物体であっても正確に把持して移載することができる。

【0099】この発明の請求項9に係る画像処理装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有するものとしたので、認識物体の積みパターンが基準となる積みパターンになっていない場合

に装置使用者や上位システムの管理者に正しく積まれていないことを警告し危険な状態を回避することができる。

【0100】この発明の請求項10に係る物体移載装置は、物体配列の基準となるデータを格納する物体配列基準データベースと、仮説生成手段が生成した物体候補組合せ仮説の妥当性を検証する仮説妥当性検証手段が認識した画像と、上記基準データと比較して正しい配列かどうかを判定する物体配列判定手段と、認識した画像が基準データに整合しないときに判定の結果を作業者に警告する警告発生手段を有する画像処理装置と移載手段とを組み合わせたものとしたので、認識物体の積みパターンが基準となる積みパターンになっていない場合に装置使用者や上位システムの管理者に正しく積まれていないことを警告し危険な状態を回避することができる。

#### 【図面の簡単な説明】

【図1】 この発明の実施の形態1の構成を示すブロック図である。

【図2】 実施の形態1の動作の流れを示すフローチャートである。

【図3】 実施の形態1における各段階において生成された画像である。

【図4】 この発明の実施の形態2の構成を示すブロック図である。

【図5】 実施の形態2の動作の流れを示すフローチャートである。

【図6】 ランダムドットパターンの照射説明図である。

【図7】 ステレオ画像のブロックマッチングの説明図である。

【図8】 この発明の実施の形態3の構成を示すブロック図である。

【図9】 実施の形態3の動作の流れを示すフローチャートである。

【図10】 低解像度距離画像から段差領域部を抽出し、高解像度距離画像の段差領域部をのみ高解像度のステレオ対応を行う状況の説明図である。

【図11】 この発明の実施の形態4の構成を示すブロック図である。

【図12】 実施の形態4の動作の流れを示すフローチャートである。

【図13】 物体候補仮説を用いて物体候補を抽出する動作状況の各段階を示す図である。

【図14】 この発明の実施の形態5の構成を示すブロック図である。

【図15】 実施の形態5の動作の流れを示すフローチャートである。

【図16】 従来の画像処理装置の処理手順を示す説明図である。

【図17】 図16に示す処理動作のフローチャートで補13-

ある。

【図18】 パターン光を用いた空間コード化の原理を示す図である。

【図19】 他の従来例の距離画像計測システムの説明図である。

【図20】 従来の別の画像処理のステレオ対応探索装置の説明図である。

【図21】 従来の別の画像処理の図20のフローチャートである。

【符号の説明】

11 画像入力手段、12 距離画像生成手段、13 最上段面領域抽出手段、14 最上段物体候補抽出手段、15 濃淡原画像格納手段、16 物体寸法データベース、17 2次元基準パターン生成手段、18 物体位置検出手段、19 情報統合手段、20 ランダム

テクスチャパターン投光手段、21 画像入力手段、22 ステレオ画像ブロック対応付手段、30 パターン投光器、31 低解像度ブロック対応付手段、32 段差領域抽出手段、33 高解像度ブロック対応付手段、34 距離画像合成手段、64 物体候補組合せ仮説生成手段、65 仮説妥当性検証手段、69 物体位置検出手段、71 物体配列基準データベース、72 物体配列判定手段、73 警告発生手段。

【手続補正2】

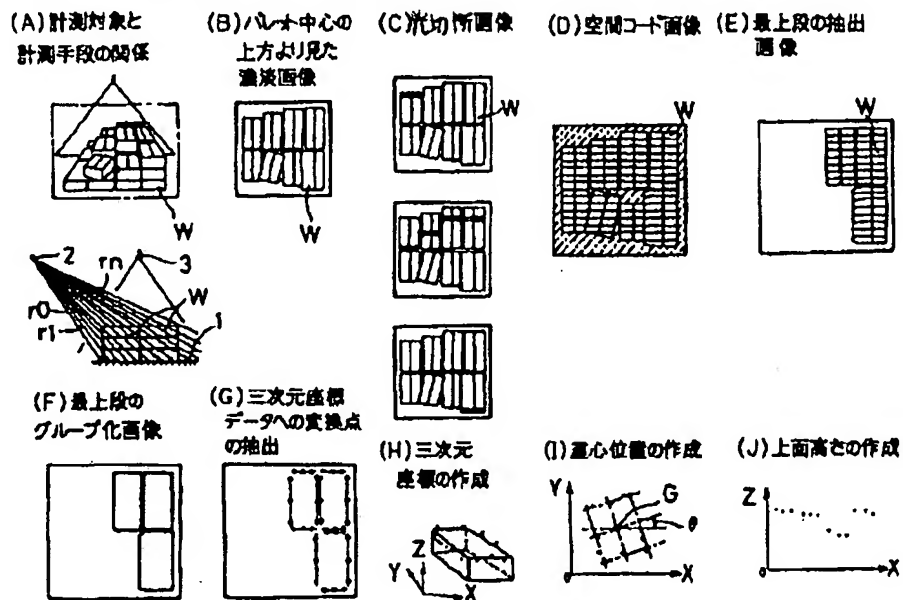
【補正対象書類名】図面

【補正対象項目名】図16

【補正方法】変更

【補正内容】

【図16】



【手続補正3】

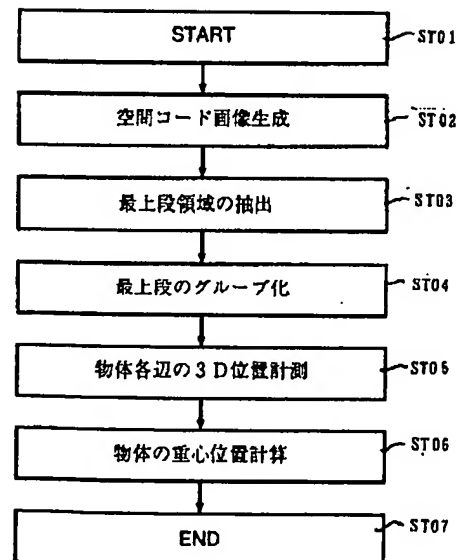
【補正対象書類名】図面

【補正対象項目名】図17

【補正方法】変更

【補正内容】

【図17】



【手続補正4】

【補正対象書類名】図面

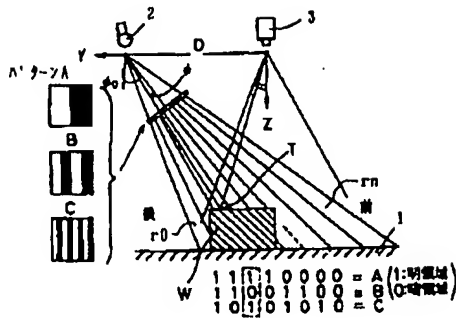


【補正対象項目名】図18

【補正方法】変更

【補正内容】

【図18】



【手続補正5】

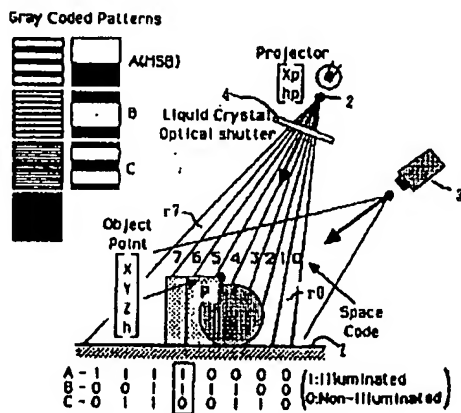
【補正対象書類名】図面

【補正対象項目名】図19

【補正方法】変更

【補正内容】

【図19】



【手続補正6】

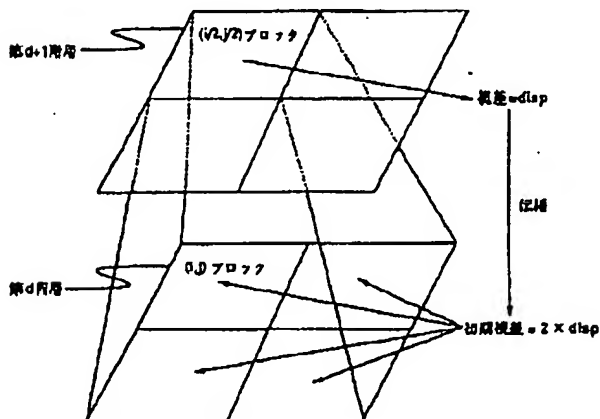
【補正対象書類名】図面

【補正対象項目名】図20

【補正方法】変更

【補正内容】

【図20】



【手続補正7】

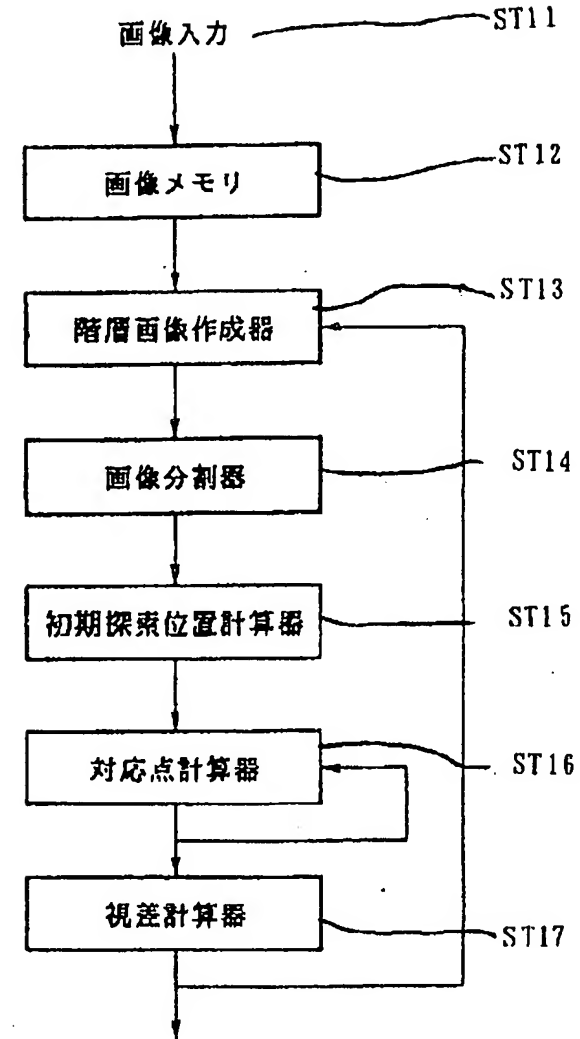
【補正対象書類名】図面

【補正対象項目名】図21

【補正方法】変更

【補正内容】

【図21】



【手続補正8】

【補正対象書類名】図面

【補正対象項目名】図22

【補正方法】削除